BIOMETRIKA

BIOMETRIKA

A JOURNAL FOR THE STATISTICAL STUDY OF BIOLOGICAL PROBLEMS

FOUNDED BY

W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

EDITED BY

EGON S. PEARSON

IN CONSULTATION WITH

HARALD CRAMÉR J. B. S. HALDANE

MAJOR GREENWOOD G. M. MORANT

JOHN WISHART

VOLUME XXIX

1937

ISSUED BY THE BIOMETRIKA OFFICE
UNIVERSITY COLLEGE, LONDON
AND PRINTED AT THE
UNIVERSITY PRESS, CAMBRIDGE

First printed in Great Britain at the University Press, Cambridge
Reprinted by offset-litho at the University Press, Oxford

CONTENTS OF VOLUME XXIX

Memoirs

Miscellanea

List of Plates and Folding Tables

Vol. XXIX. Parts I and II June, 1937

BIOMETRIKA

A JOURNAL FOR THE STATISTICAL STUDY OF BIOLOGICAL PROBLEMS

FOUNDED BY

W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

EDITED BY

EGON S. PEARSON

ASSISTED BY

GEOFFREY M. MORANT

ISSUED BY THE BIOMETRIKA OFFICE
UNIVERSITY COLLEGE, LONDON
AND PRINTED AT THE
UNIVERSITY PRESS, CAMBRIDGE

PRINTED IN GREAT BRITAIN

[Issued 8 July 1937] Reprinted by offset-litho 1952

VOLUME XXIX JUNE 1937 PARTS I AND II

BIOMETRIKA

A CONTRIBUTION TO ESKIMO CRANIOLOGY BASED ON PREVIOUSLY PUBLISHED MEASUREMENTS

BY G. M. MORANT, D.Sc.

1. *Introduction.* Among modern races of man the physical type of the Eskimos is undoubtedly one of the most specialized known. The available metrical data relating to it are far more extensive for the cranium than for any other part of the skeleton, or for the living people. Numerous studies of series of Eskimo skulls have been published in the last sixty years, but the majority of these are of little value for statistical purposes, either on account of the fact that the measurements given are inadequate, or because the series described are too small, or for both these reasons. Until recent years the monumental *Crania Groenlandica* of Professors Fürst and Hansen, published in 1915, was by far the most valuable of the contributions to the subject, but it relates only to Eastern Eskimos. The publication by Dr Hrdlička, between 1924 and 1929, of data relating to a large number of Western and Central Eskimo skulls satisfied a long-felt need. These two sets of material are the only ones dealt with, apart from comparative material for other races, in the present paper. The measurements discussed were only treated by the crudest statistical methods when first presented.

2. *Measurements of Eskimo Skulls provided by Dr Aleš Hrdlička.* Dr Hrdlička has published several contributions to Eskimo craniology, but when considering a statistical treatment of his material it is only necessary to refer to two of these:

(*a*) "Catalogue of Human Crania in the United States National Museum Collections", *Proceedings of the United States National Museum*, LXIII (1924), pp. 1–51. This, the first part of the extensive catalogue, contains individual measurements of a number of series of Eskimo skulls from different localities. There are only two of the male series sufficiently long for statistical purposes, viz. one of 40 Greenland Eskimo crania and another of 159 Eskimo crania from the north coast of St Lawrence Island in the Bering Sea. One of the former came from the Noursoak Peninsula on the west coast of Greenland, and there are no recorded localities for the other specimens in the series. No particulars regarding the discovery or age of the material are provided.

(*b*) "Anthropological Survey in Alaska", *Forty-Sixth Annual Report of the Bureau of American Ethnology*, 1928–9, pp. 19–374. This provides (pp. 254–99) a detailed discussion and mean measurements of a considerable number of groups of Eskimo crania, the majority being made up by small numbers of specimens. All the material of this kind previously described by the author is apparently

included here, and there is a considerable amount of new material. The specimens appear to be of modern or recent date, but one series discussed at greater length than the others (pp. 318–29) is believed to represent a population which lived near Point Barrow, on the north coast of Alaska, before contact with Europeans was established. This is known as the "Old Igloos" series. Dr Hrdlička says that for the purpose of this report he re-measured all the specimens which he had previously described. The means for the Greenland and St Lawrence Island series differ somewhat from those in the 1924 *Catalogue*, and the numbers on which these means are based were also changed. It is said that the individual measurements for all the material will be given in a part of the *Catalogue* which has not yet been published.

Before making statistical comparison between the different series, it is necessary to comment on the definitions followed in determining the measurements. Dr Hrdlička has published an account of his technique,* which is based on that of the Monaco Congress of 1906 with several modifications. A list of the measurements given for the Eskimo series follows, the numbers preceded by I.A. being those of the International Agreement and the letters those of the biometric technique:

(i) Maximum "glabello-occipital" length: I.A. 1. This is not precisely the same as L, defined to be the maximum calvarial length from the glabella in the median sagittal plane, but the two definitions will give readings which are either identical or very close to one another.

(ii) Maximum breadth above the mastoids and roots of zygomae: I.A. 3. This, again, is not precisely the same as B, defined to be the maximum transverse diameter on the parietal bones, but the two definitions will almost invariably give identical or closely similar readings.

(iii) Basion-bregma height: I.A. 4, *a*. Although the basion is insufficiently defined, this measurement may be supposed the same as H'.

(iv) Cranial capacity: I.A. 24, *c*. Hrdlička determined this with seed by using a method which he describes in detail. It is commonly found that different methods often give appreciably different results.

(v) Upper facial height from nasion to alveolar point: I.A. 12, $G'H$. The alveolar point is defined to be the "lowest point of the alveolar border between the two median upper incisors".

(vi) Facial length from basion to alveolar point. Hrdlička gives this definition without comment, and it may be presumed that he used the same alveolar point as in finding the upper facial height, so the measurement is GL, assuming that the basions used are the same. He diverges here from the Monaco definition

* "Anthropometry. D. Skeletal Parts: the Skull", *American Journal of Physical Anthropology*, II (1919), pp. 401–28. This article was reprinted in the author's *Anthropometry*, Philadelphia (1920). An English translation of the Monaco report is given in the same volume of the *Journal* and in the book.

(I.A. 10), which specifies that the "alveolar point" used in this case is the "median point of the anterior border of the alveolar arch", i.e. Martin's prosthion.

(vii) Chord from nasion to basion: I.A. 9. This may be supposed the same as *LB*, on the supposition that the same basions are intended.

(viii) Maximum bizygomatic breadth: I.A. 8, *J*.

(ix) Nasal height. All the definitions agree in using the nasion as the superior terminal of this measurement. According to the Monaco technique (I.A. 13) the inferior terminal is "the middle of a line connecting the lowest points of the two nasal fossae", which is inexact as the middle point will normally lie *in* the nasal spine and not on any surface. Hrdlička's practice is to "measure separately to each subnasal point and record the mean", the subnasal points being defined as "the lowest point, on each side, on the lower border of the nasal aperture, i.e. the lowest points anteriorly of the two nasal fossae". This measurement is likely to give such closely similar readings that it may be supposed the same as *NH*, *L*.

(x) Maximum breadth of nasal aperture: I.A. 14, *NB*.

(xi) Orbital breadth: I.A. 16. The terminal of this measurement nearest to the median sagittal plane is the dacryon, or "if the dacryon is obliterated, or in an abnormal situation, take the point where the posterior lacrymal crest meets the inferior border of the frontal". The lateral terminal is "the external border of the orbit, at the point where the transverse axis of the orbit meets the border, and parallel as far as possible to the superior and inferior borders". This is an inadequate definition, since there is no exact way of deciding when the dacryon is in an abnormal situation, and the point sometimes substituted for it normally gives a lesser breadth than that from the true dacryon. Hrdlička follows the Monaco instructions, and only gives data for the mean of the two orbital breadths so obtained. His measurement may be denoted by O_1' (or more precisely by $\frac{1}{2}(O_1'R + O_1'L)$), though it is not exactly the same as the true dacryal breadth.

(xii) Orbital height: I.A. 17. This is the maximum height perpendicular to the breadth and it may be supposed the same as O_2. Hrdlička gives the mean of the heights of the two orbits.

These are the only absolute measurements provided by Dr Hrdlička for Eskimo skulls which are dealt with below. He gives data for five additional measurements determined according to the Monaco definitions, viz. the length and breadth of the "upper alveolar arch" and the chord from basion to subnasal point, for which there is little comparative material; the menton-nasion height, an unreliable measurement owing to the fact that it is influenced by wear of the teeth, and the height of the mandible at the symphysis. The omission of a number of customary measurements—such as the principal arcs, minimum frontal breadth and palatal and foraminal measurements—is to be regretted. The indices and angles in curled brackets in the tables below were obtained from the mean values of the component lengths (indices) or sides of the triangle (angles) instead of from

values for individual skulls. The angles ($N\angle$, $A\angle$ and $B\angle$) are the three of the fundamental triangle of which the sides are $G'H$, GL and LB. Hrdlička gives means for the first, which he calls the facial angle, and some of our means for it differ from his quite markedly.

There is reason to suspect that he modified the ways in which some of his measurements were taken between the times when the data were obtained for the 1924 *Catalogue* and 1928 *Report*. This is suggested by a comparison of the means for the two male series given in the former year with those for the same series given after re-measurement, a number of additional specimens having been added in one case. There is a close agreement between corresponding means except in the case of the following characters:

		C	$G'H$	J	NH	$100\,NB/NH$
Greenland	1924	1560 (34)	74·4 (36)	140·0 (30)	53·4 (39)	42·9 (36)
	1928	1518 (42)	76·1 (46)	140·5 (47)	52·4 (48)	44·3 (48)
St Lawrence Island	1924	1506 (129)	76·6 (144)	140·8 (151)	55·4 (150)	44·6 (150)
	1928	1462 (142)	78·2 (139)	142·0 (148)	54·2 (148)	45·2 (148)

The differences may be partly due to the fact that the corresponding means are given for different numbers of specimens, and the sexes of some of them may have been changed, but it is impossible to avoid the conclusion that the divergences for these characters are due primarily to a change in the definitions of the measurements. It appears that no other hypothesis can explain why the facial heights increased on re-measurement while the nasal heights decreased, or why the capacities decreased while the major calvarial chords remained practically unchanged. The differences which must be attributed to personal equation are large enough to be disturbing when an attempt is made to distinguish small differences between neighbouring Eskimo types. The only means of Dr Hrdlička's series used below are those derived from the data given in his 1928 *Report*, but this contains no individual measurements and the standard deviations of two of the series in our Table I were obtained from the readings in the 1924 *Catalogue*.

3. *Measurements of Greenland Eskimo Skulls provided by Professors Carl M. Furst and F. C. C. Hansen.* The *Crania Groenlandica** of these authors is one of the most valuable and comprehensive treatises of its kind available for any race. Descriptions and detailed individual measurements of 380 crania are given, 14 of these being immature and 8 others unsexed. The sample forms a selection from the total Eskimo population of Greenland, the more densely populated west coast being represented by larger numbers of specimens than the south-west and

* *Crania Groenlandica, A Description of Greenland Eskimo Crania with an Introduction on the Geography and History of Greenland*, Copenhagen (1915).

east coasts. The distributions and means are, unfortunately, for the combined male and female series in the case of all characters except the cephalic index. A biometric treatment of the male series has been given by the present writer.* Only those constants for characters which are available for Hrdlička's series are considered below. Fürst and Hansen's orbital breadth, "from the lateral to the vertical medial margin, which is the direct continuation of the lower orbital margin and which in the Greenlanders is continued sharply and often high on the maxillary bone", may be supposed the same as the biometric orbital breadth O_1, and this is a different measurement from the dacryal breadth given by Hrdlička. There is no doubt as to whether the measurements of the two sets of data may be considered the same in definition or not except in the case of the nasal height. Fürst and Hansen do not define the way in which they determined this measurement, but it may be presumed that it was the same as, or very similar to, that employed by Hrdlička. Their cranial capacities were found with the aid of millet seed and a graduated glass cylinder, and they may also be supposed comparable to his determined by a slightly different technique.

4. *The Variabilities of Male Eskimo Series.* Standard deviations are given in Table I for the only two Eskimo series measured by Hrdlička for which individual measurements have been published (in the 1924 *Catalogue*), for Fürst and Hansen's Greenland series (the constants being taken from the paper cited), and for the long Egyptian series often used for comparative purposes.† These constants have been given for more than twice as many characters relating to the last two series. The data relate to 13 characters and the four series give 6 comparisons in pairs for each, so there is a total of 78 comparisons. In 29 cases the differences exceed three times their probable errors, and several of them are markedly significant, the highest ratio of a difference to its probable error being 9·6. Twelve of the 13 σ's for the St Lawrence Island series are less than the corresponding values for the Egyptian, and the differences exceed three times their probable errors in 8 of these cases: 11 of the σ's for Hrdlička's Greenland series are less than the Egyptian values, and the differences exceed three times their probable errors in the case of 4 of these 11 comparisons. But the σ's for Fürst and Hansen's Greenland series are in excess of the Egyptian in the case of 10 of the 13 characters, and for 4 of these 10 the differences may be supposed significant. These divergences in variability are more marked than those usually found in the comparison of cranial series believed to be racially homogeneous. This is evidently due to the fact that Hrdlička's two series show peculiarly small variation (only one character showing a significant difference in the comparison between them), while the other two show a greater and more common order of variation. It has been found

* In "Studies of Palaeolithic Man. I. The Chancelade Skull and its Relation to the modern Eskimo Skull", *Annals of Eugenics*, I (1926), pp. 257–76

† Karl Pearson and Adelaide G. Davin, "On the Biometric Constants of the Human Skull", *Biometrika*, XVI (1924), pp. 328–63.

for other material that the populations of small islands—such as the Guanche of Tenerife—are distinctly less variable than others. The distinction of Hrdlička's Greenland series in the respect considered may be due to the fact that the skulls,

TABLE I

*Standard Deviations of Male Series of Crania**

	Eskimo			Egyptian E: 26th–30th dynasties (Pearson and Davin)
	St Lawrence Island (Hrdlička)	Greenland (Hrdlička)	Greenland (Furst and Hansen)	
C	103·6 ±4·4 (129)	86·0 ±7·0 (34)	128·8 ±4·6 (175)	113·5 ±2·0 (753)
L	4·96±·19 (158)	4·36± 34 (38)	5·81±·20 (191)	5·72±·09 (895)
B	4·05±·15 (156)	4·34±·34 (36)	4·52±·16 (191)	4·76±·03 (896)
H'†	4·25±·17 (143)	3·97±·30 (39)	4·79±·17 (183)	5·03±·08 (884)
$G'H$	3·32±·13 (144)	3·96±·31 (36)	4·39±·15 (191)	4·15±·07 (845)
J	4·86±·19 (151)	5·74±·50 (30)	6·48±·23 (185)	4·57±·08 (785)
NH	2·39±·09 (150)	2·69±·21 (39)	3·10±·11 (192)	2·92±·05 (898)
NB	1·74±·07 (153)	1·67±·13 (36)	1·75±·06 (191)	1·77±·03 (893)
O_2‡	1·62±·06 (148)	1·84±·14 (38)	2·03±·07 (188)	1·88±·03 (888)
O_1'§	1·52±·06 (148)	0·92±·07 (38)	2·46±·09 (189)	1·65±·03 (880)
$100\,B/L$	2·39±·09 (156)	3·07±·25 (35)	3·00±·10 (190)	2·68±·06 (884)
$100\,NB/NH$	3·32±·13 (150)	3·74±·30 (36)	3·84±·13 (191)	3·82±·06 (881)
$100\,O_2/O_1'$‖	3·70±·15 (148)	4·83±·37 (38)	5·60±·19 (189)	4·95±·08 (876)

* The ± signs in this paper indicate probable errors, as in all earlier anthropometric papers in *Biometrika*.

† The Egyptian σ is for the vertical height from the basion (H) instead of H'. Both these measurements are available for the male Eskimo series measured by Fürst and Hansen, the σ's being 4·78±·17 for H and 4·79±·17 for H'.

‡ The σ's are for the means of the right and left orbital heights in the case of Hrdlička's series, and for the height of the left orbit in the case of the other two series

§ The σ's are for the means of the right and left orbital breadths from the dacrya (O_1') in the case of Hrdlička's series, and for the maximum breadths from the medial margin and for the left orbit only ($O_1\,L$) in the case of the other two series. Both orbital breadths have been given for a few long series and the σ's for them have been found to be in close agreement.

‖ The σ's are for the orbital indices found from the means of the heights and dacryal breadths for the right and left sides ($100\,O_2/O_1'$) in the case of Hrdlička's series, and for $100\,O_2/O_1,\,L$ in the case of the other two. These also show a close agreement when found for the same series.

of unknown origin, came entirely or in large part from a small Eskimo community. Fürst and Hansen's series may be considered a sample drawn, more or less at random, from the total Eskimo population of the country. In calculating all the coefficients of racial likeness given in this paper the Egyptian E standard deviations were used. These values are probably close to those which would be found for the majority of the series measured by Hrdlička.

5. *Comparisons of Eskimo Series by the Method of the Coefficient of Racial Likeness.* Dr Hrdlička's 1928 *Report* contains individual or mean measurements of 36 groups of male Eskimo skulls ranging in size from one to 153 individuals.

Most of the series are made up by fewer than 10 specimens, and pooling of some of the material is obviously required as a preliminary step towards statistical analysis. Three Western groups were first compiled by combining in two cases the skulls from neighbouring localities in order to make up samples of a sufficient size. The approximate positions of the sites can be seen from the map in Fig. 1, and the numbers in brackets below give the numbers of male skulls from each site. The Western groups are:

W_1—Prince William Sound (1), Kodiak Island (1), Unalaska Peninsula (1), Nushagak Bay (1), Togiak (4), Mumtrak (4), Nelson Island (9), Hooper Bay (9), Lower Yukon and delta (3), Pilot Station, lower Yukon (3), Kotlik and Pastolik (11) and St Michael Island (8).

W_2—St Lawrence Island (153).

W_3—Little Diomede Island (5), and two sites on the mainland of Asia, Indian Point (14) and Puotin (2).

The means for these three groups are given in Table II and they show a remarkably close resemblance for all characters. The crude coefficients of racial likeness are: W_1 and W_2 $-\cdot 06 \pm \cdot 22$ (18),* W_1 and W_3 $-\cdot 49 \pm \cdot 23$ (17)† and W_2 and W_3 $-\cdot 50 \pm \cdot 23$ (17)†. As far as can be seen from the data available, the Eskimo population of the south-west of Alaska, St Lawrence Island and the Asiatic mainland is perfectly homogeneous. Within this area there is no evidence of local populations differing significantly from the prevailing type, though it is quite possible that there are local variants. There is only one skull from Kodiak Island, for example, and it is quite possible that if 50 were available their measurements would distinguish the population from that of St Lawrence Island. At the moment the pooling of the three groups W_1, W_2 and W_3 appears to be justified and the combined means are those of the Western series in Table IV. It will be shown that they are not closely similar to those for any other series available.

Three groups of Eskimo skulls from the north-west and north of Alaska were made up in the following way, these being distinguished from the Western groups because their mean measurements clearly differentiate them:

NW_1—Golovnin Bay (3), Cape Nome (1), Sledge Island (5), Port Clarence (4), Wales (19), Shishmaref (13), Kotzebue (2).

NW_2—Barrow and vicinity (37).

NW_3—Point Barrow (49).

The means for these groups, given in Table II, again show a remarkably close resemblance for all characters. The crude coefficients of racial likeness between them are: NW_1 and NW_2 $-\cdot 04 \pm \cdot 23$ (17), NW_1 and NW_3 $\cdot 56 \pm \cdot 23$ (17) and NW_2 and NW_3 $-\cdot 18 \pm \cdot 23$ (17).† The pooling of the three groups again appears to be fully justified, and accordingly the combined means were computed and they

* For all the characters in Table II except GL and $B\angle$.

† For all the characters in Table II except C, GL and $B\angle$.

are those for the North-Western series in Table IV. It will be shown that they are differentiated from those for all the other series available. It is surprising to find that all the material hitherto considered in this section can be partitioned with so little hesitation into two homogeneous series which are distinctly different from one another. A crude comparison of the means in Table II shows that this

TABLE II

*Mean Measurements of Groups of Male Eskimo Skulls from Alaska**

	Western groups			North-Western groups		
	St Lawrence Island (W_2)†	South-Western (W_1)	Asiatic and Little Diomede Island (W_3)	Kotzebue Sound and *north of* Norton Sound (NW_1)	Barrow and vicinity (NW_2)	Point Barrow (NW_3)
C	1462 (142)	1503·1 (53)	1470 (5)	1448·4 (40)	—	1324 (5)
L	184·0 (153)	183·3 (55)	185·1 (21)	187·3 (47)	189·0 (37)	187·4 (49)
B	141·9 (153)	142·1 (55)	143·2 (21)	136·6 (47)	137·3 (37)	138·4 (49)
H'	136·8 (145)	135·9 (55)	137·2 (20)	138·0 (45)	137·8 (35)	137·8 (47)
LB	103·6 (145)	103·8 (54)	104·6 (19)	106·5 (45)	106·1 (35)	105·4 (47)
GL	104·3 (131)	103·7 (43)	104·4 (14)	106·0 (39)	103·9 (21)	103·9 (36)
$G'H$	78·2 (139)	78·6 (47)	78·3 (17)	77·6 (39)	78·9 (21)	78·6 (37)
J	142·0 (148)	142·1 (52)	141·9 (21)	141·8 (42)	143·4 (26)	142·6 (44)
NH	54·2 (148)	54·4 (54)	55·0 (21)	54·0 (44)	55·2 (29)	54·8 (46)
NB	24·5 (148)	24·2 (54)	25·0 (21)	23·8 (44)	23·9 (29)	23·1 (46)
O_2	36·8 (145)	36·7 (54)	37·0 (21)	36·3 (44)	36·0 (28)	36·1 (43)
O_1'	40 3 (145)	40·0 (54)	40·6 (21)	40·6 (44)	40·4 (28)	40·2 (43)
$100\ B/L$	77·1 (153)	77·6 (55)	77·4 (21)	73·0 (47)	72·6 (37)	73·9 (49)
$100\ H'/L$	{74·3 (145)}	{74·1 (55)}	{74·1 (20)}	{73·7 (45)}	{72·9 (35)}	{73·5 (47)}
$100\ B/H'$	{103·7 (145)}	{104·6 (55)}	{104·4 (20)}	{99·0 (45)}	{99·6 (35)}	{100·4 (47)}
$100\ NB/NH$	45·2 (148)	44·5 (54)	45·4 (21)	44·2 (44)	43·4 (29)	42·2 (46)
$100\ O_2/O_1'$	91·2 (145)	91·8 (54)	91·1 (21)	89·5 (44)	89·2 (28)	89·9 (43)
$N\angle$	{68°·2 (131)}	{67°·6 (43)}	{67°·8 (14)}	{68°·0 (39)}	{66°·3 (21)}	{66°·7 (36)}
$A\angle$	{67°·5 (131)}	{67°·9 (43)}	{68°·2 (14)}	{69°·2 (39)}	{69°·7 (21)}	{69°·2 (36)}
$B\angle$	{44°·3 (131)}	{44° 5 (43)}	{44°·0 (14)}	{42°·8 (39)}	{44°·0 (21)}	{44°·1 (36)}

* The indices and angles in curled brackets were derived from the means of the chords involved, instead of from values for individual skulls.

† The locations of the groups are shown on the map in Fig. 1.

step is entirely reasonable. In the case of L, B, $100\,B/L$ and $100\,B/H'$ the means for the three Western sub-groups cover a small range and those for the three North-Western sub-groups cover another small range, while there is a clear separation of the two ranges. For C, H', LB, NB, O_2, $100\,H'/L$, $100\,NB/NH$, $100\,O_2/O_1'$ and $A\angle$ the two ranges are also discrete, but the separation between them is less clear. For the remaining characters—GL, $G'H$, J, NH, O_1', $N\angle$ and $B\angle$—the two ranges overlap, but the differences are so small that all between pairs of the six groups are probably insignificant. The two major groups are thus clearly defined and clearly distinguished. It would have been expected from geographical considerations

(see Fig. 1) that the sub-group NW_1 would bear a closer resemblance to the three Western sub-groups than NW_2 and NW_3 would to these three, but there is no suggestion of this from the measurements. There appears in fact to be a distinct cleavage between the types of the south-west and north-west coast of Alaska, though more abundant material would be needed in order to ascertain with precision where the dividing line comes. Some of the series from neighbouring sites making up the sub-groups W_1 and NW_1 are so small that a few might be transferred from one to the other without affecting the position appreciably.

The means of three other series of male Eskimo skulls from Alaska are given by Dr Hrdlička in his 1928 *Report*. They fall within the same area as the Western and North-Western series dealt with above, but they were kept apart from these as their mean measurements evidently differentiate them. The first is made up by 46 specimens from Nunivak Island, the second of 131 from Point Hope and the third is the series of 27 "Old Igloos" skulls from the vicinity of Point Barrow. This last is believed to represent an earlier population than all the others, and it was kept separate on this account, and also because the type it represents is clearly distinguished from all the others determined by Alaskan series. A number of small groups from the islands west of Greenland and the Canadian mainland were pooled to form what will be called the Central Eskimo series. They are: Northern Arctic (5), Melville Peninsula (1), Southampton Island (9), Hudson Bay and Ungava Bay (5), Baffin Land, northern Devon, and vicinity (16) and Smith Sound (7).

The only remaining series measured by Dr Hrdlička is the one of 49 skulls from unknown localities in Greenland. This may be compared with Fürst and Hansen's Greenland Eskimo series. These writers conclude that: "The anthropological characters cannot contribute to a solution of the question of the migration of the Eskimos [in Greenland], owing to the fact that the homogeneity of their anthropological characters clearly shows that the Eskimos of both the west and the east coasts are of the same racial type." This conclusion is based principally on a comparison of the distributions of the measurements for *unsexed* series of crania from different regions of Greenland. Male means computed for three groups into which the total material is divided are given in Table III below in the case of nine of the more important characters. In asking whether the differences between these are significant or not, the standard deviations of Hrdlička's Greenland Eskimo series (given in Table I) were used, and it has been noticed that these are appreciably less than the values for the total series measured by Fürst and Hansen. The only differences greater than three times their probable errors are: C, Eastern less than Western ($\Delta/(\text{p.e. } \Delta) = 5{\cdot}9$) and South-Western ($6{\cdot}3$); H', Western less than South-Western ($3{\cdot}3$); J, Eastern less than South-Western ($3{\cdot}8$); NH, Western less than South-Western ($4{\cdot}0$); $100\ NB/NH$, South Western less than Western ($3{\cdot}3$). No importance can be attached to any of these differences except those for the capacity. The types appear to differ principally

in size, and it may be noted that the South-Western series has the largest means in the case of all the absolute measurements except L.

TABLE III

Mean Measurements for Male Series of Greenland Eskimo Skulls

	Western	South-Western	Eastern	Pooled	Unknown localities
	Furst and Hansen				Hrdlička
C	1536·2 (95)	1549·0 (48)	1465·8 (32)	1526·8 ±6·6 (175)*	1518 ±9·0 (42)†
L	188·6 (100)	188·5 (54)	187·9 (37)	188·4 ±·28 (191)	189·7±·42 (49)
B	134 5 (100)	135·1 (55)	133·4 (36)	134·5 ±·22 (191)	136·1±·42 (49)
H'	137 6 (97)	139·1 (51)	138·7 (35)	138·2 ±·24 (183)	139·5±·38 (49)
$G'H$	74·3 (99)	75·6 (56)	75·0 (36)	74·8 ±·21 (191)	76·1±·39 (46)
J	139·4 (96)	141·1 (53)	137·9 (36)	139·6 ±·32 (185)	140·5±·56 (47)
NH	53·4 (99)	54·6 (56)	53·5 (37)	53·75±·15 (192)	52·4±·26 (48)
$100\,B/L$	71·3 (100)	71·6 (54)	70·9 (36)	71·3 ±·15 (190)	71·8±·30 (49)
$100\,NB/NH$	43 6 (98)	42·2 (56)	42·9 (37)	43·1 ±·19 (191)	44·3±·36 (48)

* Some of the means in this column differ slightly from the values given in the paper in the *Annals of Eugenics* cited, as the latter were found from distributions instead of by direct addition.

† The probable errors in this column were found by using the standard deviations (given in Table I) for 40 skulls of the same series.

The Eskimo population of Greenland may not be quite as racially homogeneous as Fürst and Hansen supposed, but for practical purposes there can be little harm in combining the three series to form a single sample representing it. Some of the pooled means are given in Table III and they may be compared with those for Hrdlička's Greenland series in the last column of the table. It must be remembered in this case that comparison is being made between measurements taken by different people. Differences greater than three times their probable errors are only found in the case of B ($\Delta/(\text{p.e. } \Delta) = 3{\cdot}4$) and NH (4·5). There is reason to believe that the latter difference is occasioned by the fact that the nasal height was not determined in precisely the same way for the two series. Hrdlička's means for all the absolute measurements in the table, except C and NH, are greater than those for Fürst and Hansen's sample. As the facial height ($G'H$) is greater it would be anticipated that the nasal height would also be greater, but actually it is significantly the lesser. As the major calvarial chords (L, B and H') are greater for Hrdlička's than for Fürst and Hansen's series, the cranial capacity (C) would also be expected to be greater for the former, but the difference observed is of the opposite sign, though not significant. The divergences observed thus appear to be partly due to slight differences in the technique of measurement, and the size difference may be partly due to a difference in the process of sexing the skulls. In spite of these blemishes the two series of means are in close agreement. The coefficient of racial likeness between them can be computed for 16 characters

and a crude value of $\cdot 94 \pm \cdot 24$ is obtained. The highest α is for *NH* (8·2) and the next highest for *B* (5·0). It is at least possible that the coefficient differs significantly from zero merely on account of the personal equations of the measurers.

Under the circumstances, it was felt that it would be best to use the pooled means of the two series of Greenland Eskimo skulls for comparative purposes. These are given in column 3 of Table IV and the other means there are for the six series derived from Hrdlička's data, in ways described above, and finally adopted for purposes of comparing different Eskimo types with one another and with non-Eskimo types. Before treating all the Greenland Eskimo skulls as a single sample, however, it was thought advisable to make comparisons between the two sub-samples and the six series relating to Eskimo populations outside Greenland. Crude and reduced coefficients of racial likeness* between these six and the Greenland series measured by (*a*) Hrdlička, (*b*) Fürst and Hansen, and (*c*) Hrdlička and Fürst and Hansen (pooled) are given in Table V. All these coefficients differ from zero with marked significance. The first of the three series is by far the smallest and it gives the lowest crude values in all cases and values markedly lower than the others in five of the six comparisons. Series (*b*) gives intermediate values of the crude coefficients in the case of four out of the six triads, and values only slightly in excess of those for series (*c*) in the other two cases. Corresponding reduced coefficients show a much closer approach to equality, but for five of the six triads the lowest values are with series *a*, while for all six the pooled series gives intermediate values. This last relation suggests that the process of reducing the coefficients is effective, as it appears to give a measure of resemblance independent of the sizes of the samples. If the Greenland series measured by Hrdlička showed the lowest reduced coefficient in the case of comparisons with all six of the other series measured by him, this might be attributed to a difference between his technique of measurement and that employed by Fürst and Hansen. But there is one exception, in the comparisons with the "Old Igloos" series, and another explanation of the results obtained may be suggested. Nothing is known about the origin of the Greenland skulls measured by Hrdlička. If they did not form a true random sample from the total population of the country, but one biased in such a way that it bears a slightly closer resemblance to modern Western Eskimo types than this total population considered as a whole does, then its slightly lower reduced coefficients with five of the six series would be expected. It is shown below that these five are closely related to one another (see Fig. 1), and that the Greenland type does not belong to the same group though it is attached to it. The "Old Igloos" series also diverges from the Western group in the same direction as, but to a greater extent than, the Greenland series. On the hypothesis considered, Hrdlička's Eskimo series would thus be expected to be rather farther removed from the "Old Igloos" series than Fürst and Hansen's Greenland series —supposed to be a random sample from the total population of the country—is

* These coefficients are defined on pp. 100–102 of this volume of *Biometrika*.

TABLE IV

*Mean Measurements of the Male Series of Eskimo Skulls finally adopted**

	"Old Igloos": Point Barrow	Greenland‡	North-Western	Central	Nunivak Island	Point Hope	Western
C	—	1525·8 (217)	1434·6 (45)	1557·6 (17)	1504 (46)	1474 (126)	1473·1 (200)
L	192·5 (27)	188·7 (241)	187·8 (133)	189·3 (43)	188·1 (46)	184·0 (131)	183·9 (229)
B	133·0 (27)	134·7 (240)	137·5 (133)	140·2 (43)	140·9 (46)	138·6 (131)	142·1 (229)
H'	140·2 (27)	138·5 (232)	137·9 (127)	139·0 (43)	136·9 (46)	139·0 (128)	136·6 (220)
LB	107·0 (27)	105·7 (234)	106·0 (127)	106·4 (42)	105·5 (46)	104·9 (128)	103·7 (218)
GL	104·5 (20)	104·6 (229)	104·8 (96)	104·9 (37)	106·5 (42)	103·1 (105)	104·2 (188)
$G'H$	77·1 (24)†	75·1 (237)	78·3 (97)	76·9 (39)	78·3 (43)	75·2 (118)	78·3 (203)
J	141·6 (26)	139·7 (227)	142·5 (112)	143·6 (42)	143·2 (45)	143·1 (124)	142·0 (221)
NH	54·5 (27)	53·5 (240)	54·6 (119)	54·0 (43)	53·5 (44)	53·6 (126)	54·3 (223)
NB	23·7 (27)	23·0 (239)	23·6 (119)	23·0 (43)	23·5 (44)	23·9 (126)	24·5 (223)
O_2	36·2 (25)	36·3 (235)	36·2 (115)	36·2 (42)	35·9 (42)	36·3 (118)	36·8 (220)
O_1'	39·7 (25)	39·9 (47)	40·4 (115)	40·5 (42)	40·2 (42)	40·3 (118)	40·3 (220)
100 B/L	69·1 (27)	71·4 (239)	73·2 (133)	74·0 (43)	75·0 (46)	75·3 (131)	77·2 (229)
100 H'/L	{72·8 (27)}	{73·4 (232)}	{73·4 (127)}	{73·4 (43)}	{72·8 (46)}	{75·5 (128)}	{74·3 (220)}
100 B/H'	{94·9 (27)}	{97·3 (232)}	{99·7 (127)}	{100·9 (43)}	{102·9 (46)}	{99·7 (128)}	{104·2 (220)}
100 NB/NH	43·6 (27)	43·4 (238)	43·2 (119)	42·7 (43)	43·8 (44)	44·6 (126)	45·0 (223)
100 O_2/O_1'	91·3 (25)	91·4 (47)	89·6 (115)	89·2 (42)	89·2 (42)	90·1 (118)	91·3 (220)
$N\angle$	{66°·9 (20)}	{68°·1 (229)}	{67°·3 (96)}	{67°·5 (37)}	{68°·8 (42)}	{67°·4 (105)}	{68°·0 (188)}
$A\angle$	{70°·3 (20)}	{69°·9 (229)}	{69°·2 (96)}	{69°·8 (37)}	{67°·9 (42)}	{70°·2 (105)}	{67°·7 (188)}
$B\angle$	{42°·8 (20)}	{42°·0 (229)}	{43°·5 (96)}	{42°·7 (37)}	{43°·3 (42)}	{42°·4 (105)}	{44°·3 (188)}

* The indices and angles in curled brackets were derived from the means of the chords involved, instead of from values for individual skulls.

† Given in error as 261 by Hrdlička. There are 27 skulls in the series, the mean J is given for 26 and the mean 100 $G'H/J$ for 24: hence the mean $G'H$ given must be for 24 or 25 skulls.

‡ The Greenland means are the pooled values obtained from Furst and Hansen's and Hrdlička's series, and all the other series in the table were measured by Hrdlička.

TABLE V

*Coefficients of Racial Likeness between Male Greenland Eskimo Series and other Eskimo Series measured by Hrdlička**

Greenland Eskimo series measured by	"Old Igloos". Point Barrow (25·7)	North-Western (114·1)	Central (40·2)	Point Hope (123·7)	Nunivak Island (44·6)	Western (215·9)
	Crude coefficients					
Hrdlička (*H.*) (47·0)*	2·99 ± ·25 (15)	4·63 ± ·24 (16)	3 51 ± ·24 (16)	10·06 ± ·24 (16)	7·33 ± ·24 (16)	29·22 ± ·24 (16)
Fürst and Hansen (*F.*) (186·6)*	3·20 ± ·25 (15)	12·38 ± 24 (16)	8·58 ± ·24 (16)	26·30 ± ·24 (16)	16·14 ± ·24 (16)	86·30 ± ·24 (16)
H. and *F.* (pooled) (233·2)*	3·28 ± ·25 (15)	12·63 ± ·24 (16)	8·19 ± ·24 (16)	27 73 ± ·24 (16)	15·85 ± ·24 (16)	94·86 ± ·24 (16)
	Reduced coefficients					
Hrdlička (*H.*)	8·98 ± ·73	6·95 ± ·36	8 10 ± ·55	14·76 ± ·35	16·01 ± ·52	37·86 ± ·31
Fürst and Hansen (*F.*)	7·09 ± ·54	8·75 ± 17	12·97 ± ·36	17·67 ± ·16	22 41 ± ·33	43 07 ± ·12
H. and *F.* (pooled)	7·10 ± ·53	8·25 ± ·16	11·95 ± ·35	17·16 ± 15	21·16 ± 32	42·31 ± 11

* All the coefficients in this table are based on the same 16 characters, viz. C, L, B, H', LB, J, $G'H$, NH, NB, O_2, 100 B/L, 100 H'/L, 100 B/H', 100 NB/NH, $N\angle$ and $A\angle$, except those with the "Old Igloos" series for which the capacity (C) is not available. In the case of the Greenland Eskimo series the mean numbers of skulls given are for all 16 characters. These are changed slightly when C is omitted, the means for the remaining 15 being *H.* 47·3, *F.* 187 3, *H.* and *F.* 234·9.

from the "Old Igloos" series. The reduced coefficients show these relationships, but it is clear that the hypothesis cannot be justified rigorously. Only the pooled Greenland series was used in later comparisons.

Seven series were thus finally adopted: the means for these are given in Table IV and the reduced coefficients of racial likeness between them in Table VI. Fig. 1 shows the localities from which the material was obtained and the connections provided by the reduced coefficients less than 10. Three of the Alaskan and the Central series are all closely connected with one another. The Western diverges from this central group in one direction and the Greenland series diverges from it in a different direction. So far there is a general agreement between the

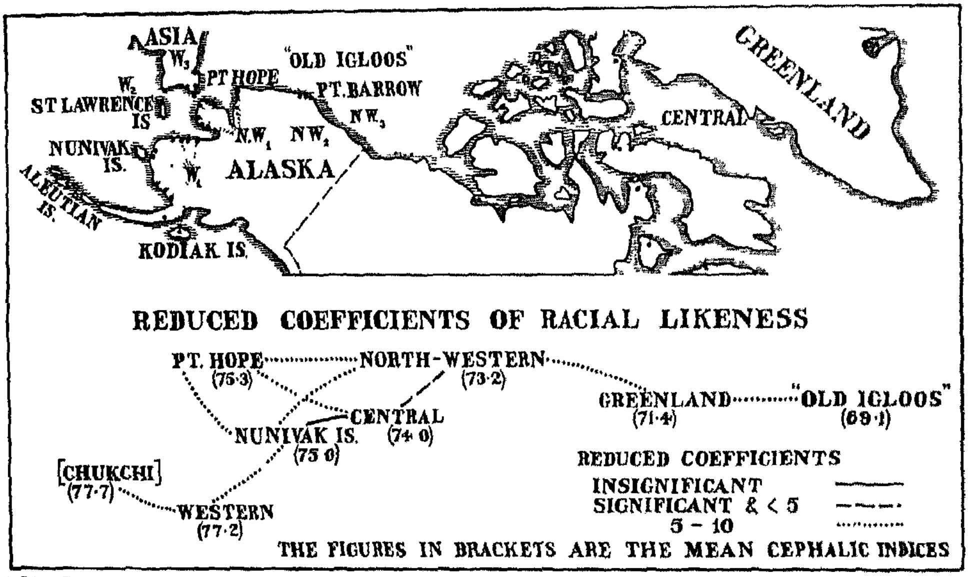


FIG. 1. THE PLACES OF ORIGIN & RELATIONSHIP OF MALE SERIES OF ESKIMO SKULLS

relationships found and the geographical positions of the populations represented, but this is not maintained in detail. The Greenland series, for example, is nearest geographically to the Central, but it bears a closer resemblance to the North-Western Alaskan than to the Central series. The extremely close resemblance of the Nunivak Island and Central Eskimo types is again unexpected. The only coefficient which differs from zero by less than three times its probable error is found in this case, but the two series show distinctly different relationships when compared with the others. It has been found in the case of other material that the fact that two series cannot be clearly differentiated when compared directly does not preclude the possibility that they will be distinguished by other comparisons. The relationships of the "Old Igloos" series have not been considered yet. This was obtained from a site in Alaska within a few miles of some of the others, but it differs from all the other Eskimo series in being older than they are, and it is also

TABLE VI

*Reduced Coefficients of Racial Likeness between Male Eskimo Cranial Series**

	$\bar{n}$	Greenland	"Old Igloos". Point Barrow	North-Western	Central	Point Hope	Nunivak Island	Western
$\bar{n}$		213·1	25·6	114·2	40·4	123·1	44 3	216 3
Greenland	213·1	—	6·10±·49 (17)	7·76±·15 (18)	11·16±·33 (18)	15·88±·14 (18)	19·44±·31 (18)	39·55±·10 (18)
"Old Igloos": Point Barrow	25 6	6·10±·49 (17)	—	16·23±·55 (17)	24·07±·73 (17)	32·91±·55 (17)	38·44±·71 (17)	67·57±·50 (17)
North-Western	114·2	7·76±·15 (18)	16·23±·55 (17)	—	2·61±·38 (18)	7·16±·19 (18)	5 49±·35 (18)	18·47±·15 (18)
Central	40·4	11·16±·33 (18)	24·07±·73 (17)	2·61±·38 (18)	—	7·00±·37 (18)	1·52±·53 (18)	17·28±·33 (18)
Point Hope	123·1	15·88±·14 (18)	32·91±·55 (17)	7·16±·19 (18)	7·00±·37 (18)	—	8 92±·35 (18)	10·93±·14 (18)
Nunivak Island	44·3	19 44±·31 (18)	38·44±·71 (17)	5·49±·35 (18)	1·52±·53 (18)	8·92±·35 (18)	—	7·11±·31 (18)
Western	216·3	39·55±·10 (18)	67·57±·50 (17)	18·47± 15 (18)	17·28±·33 (18)	10·93±·14 (18)	7·11±·31 (18)	—

* The means for the 18 characters on which these coefficients are based are given in Table IV, all the characters there being used except *GL* and *B*∠. The coefficients with the "Old Igloos" series are for 17 characters, as there is no mean capacity (*C*) available for it. The mean numbers of skulls for the characters used ($\bar{n}$'s) are for all 18 in the case of all the series except the "Old Igloos". In the comparison of this with the six others their $\bar{n}$'s are slightly different from the values given for all 18 characters.

the smallest series used. The type shows no affinities to any of the modern ones found in the same area and it only shows one close connection, which is with the Greenland series.

The material available shows that there are a number of existing Eskimo populations of types which can be clearly distinguished from one another, and there is evidence of one type which appears to be extinct to-day. The fact that no close correlation is found between geographical position and resemblance as measured by the reduced coefficient of racial likeness may be due to migrations of the peoples represented. Evidence of other extinct Eskimo populations will probably be needed in order to throw light on the origin of the present-day varieties, but comparison with non-Eskimo material should also aid a solution of this problem.

6. *Comparisons of Single Characters.* The means of the seven Eskimo series finally adopted for purposes of racial comparison are given in Table IV. They relate to 20 characters and the coefficients of racial likeness were computed for 18 of these. In the process of computation a value, α, is obtained for each character in each comparison; this is approximately the square of the ratio of the difference between two means to its standard error. For the seven series there are $7 \times 6/2 = 21$ comparisons for each character, except in the case of the capacity (C), for which the total is 15, as one mean is missing. We may decide, quite arbitrarily, to consider that an α indicates a significant difference if it is greater than 10. If samples were drawn from two populations which actually had identical means for a particular character, then an α greater than 10 would only be expected to occur once in about 625 trials. The numbers of α's greater than 10 in the 21 comparisons will give estimates of the relative degrees to which the coefficients are determined by different characters. As is usually found in such comparisons, there are marked distinctions between the characters when examined in this way, some being practically constant for all the series and others showing significant differences between most pairs of them.

The three orbital measurements and the nasal angle show no α's at all greater than 10. The nasal height is almost as constant, as it only shows one value greater than the limit chosen and this is only slightly above the limit: the α is 11·30 for the Greenland and North-Western series. A second group of characters may be distinguished by the fact that they only show significant differences in some of the comparisons between the Western series on the one hand, and the other six series on the other. These are the nasal index (3 α's > 10), the basio-bregmatic height (3 α's > 10) and the chord from nasion to basion (LB: 4 α's > 10). It can be seen from Table IV that the Western series has the highest mean for $100\,NB/NH$ and the lowest means for H' and LB. Two other characters may be added to this group. The alveolar angle ($A\angle$) show 7 of the 21 α's greater than 10, and 5 of these —including the only two α's indicating markedly significant differences—are for

comparisons between the Western and the other series: the Western has the lowest mean. The nasal breadth shows 5 α's greater than 10, and 4 of these are for comparisons between the Western and the other series: the Western has the highest mean. This disposes of 10 of the 18 characters: 5 of these 10 may be supposed practically constant for all the Eskimo types, and the other 5 characters are practically constant for 6 series, but they distinguish these 6 from the Western Eskimo series.

The remaining 8 characters serve other purposes The bizygomatic breadth (J) only shows 5 of its 21 α's significant, though all these indicate clear divergence, and they are all for comparisons between the Greenland Eskimo and the other series: the Greenland mean is not distinguished from the "Old Igloos", but it is from all the others. The height-length index again distinguishes one series from all the others, but in this case it is the Point Hope: all its α's for the character indicate clear significance, and there is only one other α greater than 10 (viz. 10·52). The significant differences are more erratic in the case of the upper facial height ($G'H$: 6 α's out of 21 greater than 10) and of the capacity (C: 5 α's out of 15 greater than 10). The remaining 4 characters are distinguished from all the others by the fact that they show more significant than insignificant differences. In each one of these cases there are 21 comparisons and the numbers of α's greater than 10 are 13 for L, 14 for B, 15 for $100\,B/H'$ and 17 for $100\,B/L$.

In the comparison of a group of series representing closely related populations, it is commonly found that the major calvarial chords and the indices derived from them show a larger percentage of significant differences than any other characters, and the cephalic index generally distinguishes the types more effectively, on the average, than the calvarial length or breadth from which it is derived. In Table IV the series are arranged in order of their mean cephalic indices: L, B and $100\,B/H'$ give very similar orders to this, but the same is not true for any other character. By considering the characters singly and then attempting to combine the evidence of each, it does not seem to be possible to construct any clear picture of the situation. Different characters suggest different conclusions and the advantage of using a generalized criterion, such as the coefficient of racial likeness, is evident.

The coefficients (Fig. 1) suggest that four of the Eskimo series represent populations which are all closely related to one another. The Greenland Eskimos diverge from this central group in one direction, and the type known from the "Old Igloos" skulls diverges in the same direction, but to a greater extent: the Western Eskimos diverge from the central group in another direction. This arrangement is suggested by the calvarial breadth (B) and two indices ($100\,B/L$ and $100\,B/H'$) which involve this measurement. The same arrangement is not suggested by any other character for which means are given in Table IV, but it is by another index involving B, viz. the "cranio-facial" ($100\,J/B$). Means for this, computed not from individual measurements but from the means of the two

chords, are: "Old Igloos" 106·5 (26), Greenland 103·7 (227), North-Western 103·6 (112), Central 102·4 (42), Nunivak Island 101·6 (45), Point Hope 103·2 (124) and Western 99·9 (221). Dr von Bonin has given* means of this index for 49 male cranial series and the highest value in his list is 104·9 for Loyalty Islanders.

The mean calvarial length of the "Old Igloos" series (192·5) is almost as great as the largest recorded for any male series of skulls; the nasal breadth of 23·0 (Greenland and Central series) is very close to the extreme found for all races, and the nasal index of 42·7 (Central) appears to be the lowest as yet recorded. These, however, are not characters which distinguish the "Old Igloos" and Greenland from all the other Eskimo types.

It may be noted that the Eskimo skull also appears to be quite extreme among modern races of man in having the "flattest" facial skeleton, though its nasal bridge is not peculiarly flat.† It has also been shown that its malar bones are extremely large and that an index expressing their vertical arcs as percentages of their horizontal arcs makes a clear distinction between the Eskimo and all other races for which data are available.‡ These measurements have only been given for a single series of Eskimo skulls—viz. one made up principally by specimens from Greenland—and their means for the series measured by Dr Hrdlička should be of particular interest. Other features which cannot be estimated from any measurements available, such as the median sagittal crest, also demonstrate that the Eskimo type is peculiarly specialized.

7. *Comparisons between Eskimo and Asiatic Series.* As a preliminary to any discussion of the "origin" of the Eskimo, it is clear that comparisons must be made between the different varieties found and series representing other races. The type is certainly one of the most specialized known, and the fact that several distinct varieties of it are found should aid the solution of problems concerning its relationships. In spite of the striking resemblance of the Chancelade to modern Eskimo skulls, there is no race known to have existed in Europe since palaeolithic times which is closely similar to that of the northern people. It is to be expected, however, that close affinities will be found with certain American and Asiatic races. It is hoped that the results of statistical comparisons between the Eskimo and North American cranial material will be presented later,§ and only comparisons with Asiatic material are considered here.

Coefficients of racial likeness for all pairs of 26 male Asiatic series have been published.‖ Their full comparison by the same method with the seven Eskimo

* *Biometrika*, XXVIII (1936), p. 133.

† See T. L. Woo and G. M. Morant, "A Biometric Study of the 'Flatness' of the Facial Skeleton in Man", *ibid.* XXVI (1934), pp. 196–250.

‡ See T. L. Woo, "A Biometric Study of the Human Malar Bone", *ibid.* XXIX (1937), pp. 113–23.

§ In a paper by Dr von Bonin and the writer which is nearly completed.

‖ T. L. Woo and G. M. Morant, "A Preliminary Classification of Asiatic Races based on Cranial Measurements", *Biometrika*, XXIV (1932), pp. 108–34. The Tibetan *B* series of 15 skulls was omitted because it is too short for the purpose in view.

series would necessitate the computation of 182 coefficients. But it has been decided that for purposes of classification the lower orders of reduced coefficients only should be used, and it is to be anticipated that the vast majority of these 182 will be of higher orders which can be neglected. The method described below makes it possible to decide, from a comparison of the means for a few characters, whether the sets of means for two series may possibly give a reduced coefficient less than a particular value (19), or whether it will be safe to assume that the coefficient will be greater than this limit. If the simple test indicates the second of these conclusions, then there is no need to calculate the coefficient, as it will not be needed in the classification. The method has been used in the comparison of other groups of series, and it makes it unnecessary to carry out a large amount of computation.

In classifying the 26 Asiatic series, all reduced coefficients less than 19 were neglected. It was found for the total 325 ($=26 \times 25/2$) comparisons that the calvarial length, breadth and height and the three indices derived from these chords gave numbers of significant α's larger than, or almost as large as, the numbers given by any other of the 31 characters used The values of the coefficients are evidently determined in large part by these six measurements, though others also play important rôles. The maximum differences between the means found in the case of the 54 comparisons giving reduced coefficients less than 19 are:

L	B	H'	$100\,B/L$	$100\,H'/L$	$100\,B/H'$
6·7 mm.	6·1 mm.	6·3 mm.	5·4	3·4	6·5

These values are much less, of course, than the corresponding maximum differences which would be found in the case of all possible comparisons between pairs of the 26 Asiatic series. If any one of these series could be compared with a new Asiatic series, and if any one of the differences of the means for the six characters were found to be greater than the value for the character given above, then it is unlikely that the reduced coefficient found in this case would be less than 19. Under the same circumstances, it is still less likely that one of the Asiatic and a non-Asiatic series would give a reduced coefficient less than 19. These considerations can be used to select those pairs of series in new comparisons which will be the only ones likely to provide reduced coefficients less than the limit which has been arbitrarily chosen. The ranges of the differences actually used for this purpose were those above with the addition of ·1 to each, viz. L 6·8 mm., B 6·2 mm., H' 6·4 mm., $100\,B/L$ 5·5, $100\,H'/L$ 3·5 and $100\,B/H'$ 6·6.

Comparisons of means restricted to these six characters were first made between each of the seven Eskimo series, on the one hand, and each of the 26 Asiatic series, on the other. If for a particular pair the difference found between the means was found to be in excess of the limit fixed in the case of any one or more of the characters then no calculation was carried out, as it may be presumed that all such pairs would give reduced coefficients of racial likeness greater than

19. This speedy test made it unnecessary to calculate 161 of the total 182 coefficients possible. The remaining 21* might give connections of the order finally considered, but the majority of them would still be expected to be of a higher order. In each of these cases the α's were first calculated for characters which showed most significant differences, and it was generally possible to see from a few of these that the reduced coefficient must exceed 19. It was only necessary to calculate two coefficients in full, and one of these was found to exceed 19 while the other is:

Western Eskimo ($\bar{n} = 220 \cdot 8$) and Chukchi (34·1)—reduced coefficient $= 7 \cdot 05 \pm \cdot 45$ for 13 characters.

In this comparison the nasal index ($\alpha = 14 \cdot 0$) is the only character which gives an α greater than 10. The Chukchi series (measured by Fridolin†) represents a people, inhabiting the extreme north-east of Asia, generally supposed to have close physical affinities to the Eskimos. It only showed one reduced coefficient less than 19 with the other Asiatic series, viz. that of $18 \cdot 27 \pm \cdot 65$ with the Prehistoric Chinese. The modern Chinese can thus be linked to the Greenland Eskimo type by a number of intermediate types, the sequence being: Modern Chinese—Prehistoric Chinese—Chukchi—Western Eskimo—Central Eskimo—Greenland Eskimo.

* The test also allows four comparisons between the Tibetan *B* and the Eskimo series, but all the reduced coefficients were found to be greater than 19.

† The calvarial height given for the Chukchi skulls is the vertical from the basion (H) in place of the more usual basio-bregmatic (H'). One mm. was subtracted from the mean H to give an approximation to H', as average differences very close to this have been found for all the longest series for which both heights have been given.

ON THE z-TEST IN RANDOMIZED BLOCKS AND LATIN SQUARES

By B. L. WELCH, Ph.D.

CONTENTS

1. *Introduction.* The validity of Fisher's z-test in the practical situations in which it has been applied has been the subject of much discussion.* In general, the mathematical distribution of z follows from the assumption that the sample observations x_i $(i=1, 2, \ldots, n)$ can be written

$$x_i = c_{i1}\theta_1 + c_{i2}\theta_2 + c_{i3}\theta_3 \ldots c_{ir}\theta_r + \eta_i, \qquad \ldots\ldots(1)$$

where the c's are known numbers, the θ's are $r<n$ unknown parameters, and the η's are normally and independently distributed about zero with standard deviations proportional to known numbers. Results following from such a starting-point may be termed results from normal theory. In practice we may not wish to make all the above assumptions regarding the η's, and to a certain extent it can be shown that, not doing so, we can still use the tests based upon them. Of especial interest are the cases of experimentation into which randomization enters as part of the structure. R. A. Fisher has pointed out† that, in any such case, it is possible to carry through arithmetical calculations, from which the hypothesis under test may be judged, without making any assumptions whatever. These calculations are lengthy. One can, however, consider only certain

* For a bibliography of the subject see a paper by T. Eden and F. Yates entitled "On the Validity of Fisher's z-test when Applied to an Actual Example of Non-Normal Data", *J. agric Sci.* XXIII (1933), pp. 6–16. Other references are given later.

† See, for instance, *The Design of Experiments*, Oliver and Boyd (1935), p. 51.

aspects of them, which are sufficient to give useful comparisons with the results from normal theory. This I have done in the present paper.

2. *Randomized Blocks.* The first situation to be discussed will be that of Randomized Blocks. Here the treatments under test are represented once and once only in each of a number of blocks. The yields given by an experiment may be denoted by $x_{i(k)}$, where $i\,(=1, 2, \ldots, n)$ are the blocks and $k\,(=1, 2, \ldots, s)$ the treatments. The usual procedure employed to test whether the treatments can be regarded as equivalent is to perform on the yields the analysis of Table I and

TABLE I

Analysis of Variance for Randomized Blocks

Source	Degrees of Freedom	Sum of Squares	Mean Square
Between Treatments	$f_1=(s-1)$	S_1	$v_1=S_1/f_1$
Between Blocks	$f_2=(n-1)$	S_2	$v_2=S_2/f_2$
Residual	$f_0=(n-1)(s-1)$	S_0	$v_0=S_0/f_0$
Total	$(ns-1)$	S	

to calculate the criterion $z=\frac{1}{2}\log_e(v_1/v_0)$. This criterion is then referred to a certain theoretical distribution, namely the distribution of z obtained by assuming that

$$x_{i(k)}=A_i+\eta_{ik}, \qquad \ldots\ldots(2)$$

where the A's are unknown block means and the η's are all normally and independently distributed about zero, with the same unknown standard deviation.

To investigate the meaning and extent of these assumptions it is necessary to consider further details of the experimental arrangement, and the exact manner in which the hypothesis, that the treatments are equivalent, is usually formulated. For convenience let the plots in each block be numbered $j=1, 2, \ldots, s$. Then the yield which the kth treatment would give, if applied under the experimental conditions, on the jth plot of the ith block may be denoted by $x_{ij(k)}$. For any plot (i, j), of course, only one of the quantities $x_{ij(k)}$ is real, viz. the one for the treatment k which is actually used on that plot in the experiment. The other $x_{ij(k)}$ are hypothetical, based on the conception of what might happen if the experiment could be repeated under the same essential conditions, using in turn every treatment on the plot (i, j). The hypothesis that the experiment is to test must, for statistical purposes, be expressed as a relation holding in some hypothetical population. In this case the population consists of the values $x_{ij(k)}$.* In the

* For a further discussion of this manner of defining our statistical population see the concluding section of the paper.

literature the hypothesis has been formulated in terms of these x's, in two distinct ways.

(i) R. A. Fisher has considered the hypothesis that the treatments would give equivalent results on *every individual plot*, i.e. that $x_{ij(k)}$ would be the same for all k.

(ii) J. Neyman has suggested that we should allow the possibility that the treatments would affect individual plots differentially, and should consider the hypothesis that the average yield of each treatment, if applied over the *whole experimental field*, would be the same. This means that $x_{..(k)}$ would be the same for all k, where $x_{..(k)}$ is the mean of $x_{ij(k)}$ over all (i, j).

The first of these "null"* hypotheses is the one which will be considered here. We may express it by the following equation:

$$x_{ij(k)} = x_{ij} \qquad (k = 1, 2, \ldots, s). \qquad \ldots\ldots(3)$$

The other "null" hypothesis was discussed in a paper given recently by Dr Neyman to the Royal Statistical Society,† in relation to the same problem that I am concerned with here. He, too, investigated the influence on the z-test, of the fact that the assumptions of normality and independence in equation (2) are not exactly satisfied. He came to certain conclusions and expressed the opinion that further investigation was desirable. The results that I obtain in the present paper may, I think, profitably be compared with his. I should emphasize, however, that the "null" hypothesis which I am using is that of equation (3) and that the situations are therefore not exactly the same. However, as Dr Neyman points out in commenting on the discussion after his paper, his results are applicable to the "null" hypothesis of (3), if some of the quantities in his equations are given certain values. On the other hand the methods I adopt here are not applicable to his more general "null" hypothesis.

We must now refer to the essential point of the arrangement of Randomized Block experiments. This is as follows. In every block the s treatments are assigned entirely at random to the s plots available for them. This means that, if the hypothesis of equivalent treatments is true (i.e. if (3) is satisfied), the yields $x_{i(k)}$ $(k = 1, 2, \ldots, s)$, given by the experiment will be a random arrangement of x_{ij} $(j = 1, 2, \ldots, s)$. For instance, in Fig. 1 (a) is given a possible set of yields x_{ij} for a field consisting of four blocks with three plots each. Fig. 1 (b) shows one possible way in which the treatments may be arranged on this field. In the first block, treatments 1, 2 and 3 are on plots 2, 1 and 3, respectively; this is only one of $3!$ possible arrangements. Similarly, in the other three blocks we have illustrated one of the $3!$ possible arrangements. Hence, taken as a whole, Fig. 1 (b) represents one of $(3!)^4$ possible arrangements.

* The term "null hypothesis" is used in the literature to denote the hypothesis that the treatments are equivalent

† "Statistical Problems in Agricultural Experimentation", *J. Roy. statist. Soc.* Suppl. II No. 2 (1935), pp. 107–180.

The method of randomization gives all the possible arrangements an equal chance of occurring. Corresponding to each arrangement there will be a different reclassification of the yields by block and treatment. Fig. 1 (*c*), for instance, shows the reclassification of the yields of Fig. 1 (*a*) when the arrangement Fig. 1 (*b*) is applied to the field. From the reclassified yields the analysis of variance of Table I can be worked out and the value of z computed. The $(3!)^4$ arrangements will each lead to a value of z, and the z of the experiment may be regarded as randomly selected from this distribution of values. The question whether the theoretical z-distribution of normal theory gives a valid test of the hypothesis of treatment equivalence, therefore, involves a comparison of the theoretical distribution of z with the distribution which would be obtained by taking all the possible arrangements.*

Blocks (i)	Numbers of Plots (j) 1	2	3
1	16	12	10
2	26	22	24
3	14	11	6
4	31	37	40

Fig. 1 (*a*). Example of possible yields x_{ij} on 4 block by 3 plot field.

Blocks (i)	Numbers of Plots (j) 1	2	3
1	2	1	3
2	3	1	2
3	1	3	2
4	2	3	1

Fig. 1 (*b*). Possible arrangement of treatments k $(=1, 2, 3)$ on the field of Fig. 1 (*a*).

Blocks (i)	Treatments (k) 1	2	3
1	12	16	10
2	22	24	26
3	14	6	11
4	40	31	37

Fig. 1 (*c*). Reclassification of yields $x_{i(k)}$ obtained by applying the arrangement Fig. 1 (*b*) to the field Fig. 1 (*a*).

3. *Normal Theory and Randomization Compared.* One approach to the comparison of the z-distribution from normal theory with that from randomization is to take separately the mean squares v_0 and v_1 of which z is a function. In normal theory v_0 and v_1 are independently distributed and their mean values are both σ^2, where σ is the standard deviation of a single η in equation (2). From randomization it is found that v_0 and v_1 have equal expectations. They are not, however, independent, since the sum (S_0+S_1) is constant, being the total sum of squares within blocks and therefore not dependent on the manner in which the treatments are assigned within blocks. The parallelism between the two theories also breaks down if we consider the variances of v_0 and v_1.

In the normal theory v_0 and v_1 are distributed as $(\chi_0^2\sigma^2)/f_0$ and $(\chi_1^2\sigma^2)/f_1$ respectively, where χ_0^2 and χ_1^2 are independent χ^2's with f_0 and f_1 degrees of freedom. The variances of v_0 and v_1 are therefore $2\sigma^4/(n-1)(s-1)$ and $2\sigma^4/(s-1)$, i.e. are in the ratio of $1:(n-1)$. From randomization, since (S_0+S_1) is constant, the variance of S_0 must be the same as that of S_1. v_0 and v_1 therefore have variances

* For brevity, any frequency constants calculated over all the possible random arrangements will be termed constants calculated from *randomization*. It should be emphasized that we are only considering what happens if the hypothesis of equation (3) is really true.

proportional to $1/f_0^2$ and $1/f_1^2$, i e. variances in the ratio $1:(n-1)^2$. With such discrepancies between results from normal theory and from randomization, it would be unsafe to conclude, from the circumstance that the expectations of v_0 and v_1 are in agreement, that the z-distributions will also compare favourably. It is advisable to try to obtain some results for the z-distribution directly,* and this I have done by means of a simple transformation.

It is of interest to note here a sampling experiment which was designed to compare directly the z-distribution from randomization with that from normal theory, for one particular set of data. This practical investigation, carried out by T. Eden and F. Yates,† did show close agreement between the z-distributions obtained in the two ways. We shall return to their example later.

Instead of z, we can equally well use the function of z:

$$U = \frac{S_1}{S_0+S_1} = (\overline{n-1}\, e^{-2z}+1)^{-1}, \qquad \text{......(4)}$$

which increases monotonically with z. Instead of saying that we reject the hypothesis of equivalent treatments when $z > z_0$ (say), we shall now say that we reject when $U > U_0$, where U_0 is related to z_0 by equation (4). If the U distributions from normal theory and from randomization compare favourably, then necessarily the z-distributions will do so also. The convenience of U lies in the fact that in the randomization procedure (S_0+S_1) is constant, and thus only the variation of S_1 need be considered.

The comparison of the U distributions will be made through the medium of their first two moments. In the normal theory we have

$$U = \frac{(\chi_1^2\sigma^2)}{(\chi_0^2\sigma^2+\chi_1^2\sigma^2)} = \frac{\chi_1^2}{(\chi_0^2+\chi_1^2)},$$

where χ_0^2 and χ_1^2 are independently distributed as χ^2 with degrees of freedom $f_0=(n-1)(s-1)$ and $f_1=(s-1)$, respectively. It follows that the distribution of U is

$$p(U) = \text{const.} \times U^{(f_1/2)-1}(1-U)^{(f_0/2)-1}. \qquad \text{......(5)}$$

The moments of this distribution are

$$\left\{\begin{aligned} &\text{mean}\, U_N = \frac{f_1}{f_0+f_1} = \frac{1}{n}, \\ &{}_N\mu_2' = \frac{f_1(f_1+2)}{(f_0+f_1)(f_0+f_1+2)} = \frac{s+1}{n(ns-n+2)}, && \text{......(6)} \\ &\sigma_{U_N}^2 = \frac{2(n-1)}{n^2(ns-n+2)}. && \text{......(7)} \end{aligned}\right.$$

* Dr Neyman in the discussion after his paper already referred to, pointed out this advisability of considering the z-distribution directly, when any investigation of the validity of the z-test is being made.

† "On the Validity of Fisher's z-test", *loc. cit.*

The suffix N is here used to denote that the moments are calculated from the normal theory. The suffix R will be used for the moments of U from randomization. The calculation of these moments is made easier by denoting the means of the yields x_{ij} over the blocks by B_i, and the deviations from these means by u_{ij}. Thus

$$x_{ij} = B_i + u_{ij}, \qquad \text{......(8)}$$

where $\sum_j u_{ij}$ is zero. The experimental yields $x_{i(k)}$ will then be written

$$x_{i(k)} = B_i + y_{i(k)}. \qquad \text{......(9)}$$

Using the dot notation to indicate that a mean is being taken over all values of the letter replaced by the dot, we have

$$(S_0 + S_1) = \sum_i \sum_k (x_{i(k)} - x_{i(\cdot)})^2 = \sum_i \sum_k y^2_{i(k)},$$

since $x_{i(\cdot)}$ must equal B_i. Also, since $x_{i(k)}$ $(k = 1, 2, \ldots, s)$ are a random arrangement of x_{ij} $(j = 1, 2, \ldots, s)$, $y_{i(k)}$ must be a random arrangement of u_{ij}. Hence

$$(S_0 + S_1) = \sum_i \sum_j u^2_{ij}. \qquad \text{......(10)}$$

Also

$$S_1 = \sum_k n\,(x_{\cdot(k)} - x_{\cdot(\cdot)})^2 = \sum_k n y^2_{\cdot(k)} = \frac{1}{n} \sum_k \{(\sum_i y_{i(k)})^2\}$$

$$= \frac{1}{n} \sum_k \{\sum_i y^2_{i(k)} + \sum_{i \neq m} y_{i(k)} y_{m(k)}\}^*$$

$$= \frac{1}{n} \sum_i \sum_j u^2_{ij} + \frac{1}{n} \sum_{i \neq m} \sum_k y_{i(k)} y_{m(k)}.$$

Since $\sum_j u_{ij}$ is zero, the expectation of any $y_{i(k)}$ is zero. Further, since the arrangements in different blocks are made independently, the y's in different blocks are independent. Hence, using E to denote expectation, we have

$$E\,(S_1) = \frac{1}{n} \sum_i \sum_j u^2_{ij}. \qquad \text{......(11)}$$

Hence

$$\text{mean } U_R = \frac{E\,(S_1)}{(S_0 + S_1)} = \frac{1}{n}, \qquad \text{......(12)}$$

which is the same as the mean from normal theory. For the variance of U_R, first consider

$$S_1^2 = \frac{1}{n^2} [\sum_k \{\sum_i y_{i(k)}\}^2]^2,$$

i.e.

$$n^2 S_1^2 = \sum_k \sum_{k'} \sum_i \sum_m \sum_p \sum_q y_{i(k)} y_{m(k)} y_{p(k')} y_{q(k')}.$$

This summation is taken for k and k' over $1, 2, \ldots, s$, and for i, m, p, and q over $1, 2, \ldots, n$. There are thus $n^4 s^2$ terms, but not all of these contribute to the expecta-

* Single summations are over all values of the letter indicated. $\sum_{i \neq m}$ indicates a summation over all values of i and m excluding $i = m$, i.e. the summation includes both $y_{1(k)} y_{2(k)}$ and $y_{2(k)} y_{1(k)}$, the fact that these terms are the same being ignored. This convention is used throughout.

tion of $n^2 S_1^2$. For if, in any term, any one of the subscripts i, m, p and q is not equal to some one of the other three, the expectation of that term must be zero. Hence

$$\begin{aligned} n^2 E[S_1^2] &= E[\sum_k \sum_{k'} \{\sum_i y_{i(k)}^2 y_{i(k')}^2 + \sum_{i \neq m} y_{i(k)}^2 y_{m(k')}^2 + 2 \sum_{i \neq m} y_{i(k)} y_{m(k)} y_{i(k')} y_{m(k')}\}] \\ &= E[\sum_k \sum_{k'} \{(\sum_i y_{i(k)}^2)(\sum_i y_{i(k')}^2) + 2 \sum_{i \neq m} y_{i(k)} y_{m(k)} y_{i(k')} y_{m(k')}\}] \\ &= E[\{\sum_k \sum_i y_{i(k)}^2\}^2] + 2 \sum_k \sum_{k'} \sum_{i \neq m} E[y_{i(k)} y_{i(k')}] E[y_{m(k)} y_{m(k')}] \\ &= (\sum_i \sum_j u_{ij}^2)^2 + 2s \sum_{i \neq m} E[y_{i(k)}^2] E[y_{m(k)}^2] \\ &\qquad + 2s(s-1) \sum_{i \neq m} E[y_{i(k)} y_{i(k')}] E[y_{m(k)} y_{m(k')}], \qquad \text{......(13)} \end{aligned}$$

k and k' in the last term standing for any two different treatments.

Now $E[y_{i(k)}^2] = (\sum_j u_{ij}^2)/s$, and for $k \neq k'$ we have

$$E[y_{i(k)} y_{i(k')}] = \frac{\sum_{j \neq j'} u_{ij} u_{ij'}}{s(s-1)} = \frac{-\sum_j u_{ij}^2}{s(s-1)}.$$

Hence (13) gives

$$\begin{aligned} n^2 E(S_1^2) &= (\sum_i \sum_j u_{ij}^2)^2 + \frac{2}{s-1} \sum_{i \neq m} (\sum_j u_{ij}^2)(\sum_j u_{mj}^2) \\ &= \frac{(s+1)(\sum_i \sum_j u_{ij}^2)^2 - 2\sum_i (\sum_j u_{ij}^2)^2}{(s-1)}. \end{aligned}$$

Hence from (10),

$$E(U_R^2) = \frac{(s+1) - 2A}{n^2(s-1)}, \qquad \text{......(14)}$$

where

$$A = \frac{\sum_i (\sum_j u_{ij}^2)^2}{(\sum_i \sum_j u_{ij}^2)^2}. \qquad \text{......(15)}$$

Since the mean U_R is $1/n$, we get from (14),

$$\sigma_{U_R}^2 = \frac{2(1-A)}{n^2(s-1)}. \qquad \text{......(16)}$$

As the mean U is the same from normal theory and from randomization, a comparison of the two U distributions can be made, to a first approximation, from the variances of equations (7) and (16). In this discussion attention will be focussed on what has been termed by Neyman and Pearson *the first kind of error*, i.e. the risk of rejecting the hypothesis that the treatments are equivalent when it is actually true, as distinct from *the second kind of error* which occurs when we fail to detect differentiation where it really exists. Suppose we wish the risk of the first kind of error to be ϵ. Let U_0 be the value from normal theory such that $P(U_N > U_0) = \epsilon$. Then the rule adopted to test whether the experimental results are consistent with there being no real differences in treatments, is to reject this

hypothesis if the U of the experiment is greater than U_0. The real chance of the first kind of error should not, however, be drawn from the U_N distribution but from the U_R distribution. Therefore we need to know how $P(U_R > U_0)$ compares with ϵ. Roughly, we may say that if $\sigma_{U_R} < \sigma_{U_N}$ the chance of $U_R > U_0$ will be less than ϵ: on the other hand if $\sigma_{U_R} > \sigma_{U_N}$ the chance of $U_R > U_0$ will be greater than ϵ.

From equation (16) it is seen that apart from n and s, the comparison of σ_{U_N} and σ_{U_R} involves only the single function A of the plot yields. This function depends on the relative sizes of $(\sum_j u_{ij}^2)$ in the different blocks, i.e. on the relative sizes of the observed variances within the blocks. Now the minimum value A can have is when $(\sum_j u_{ij}^2)$ is the same for each block. A is then equal to $1/n$. Further, since each $(\sum_j u_{ij}^2)$ is essentially positive, it is seen that the maximum value A can have is when $(\sum_j u_{ij}^2)$ is zero in every block, except one. A is then equal to unity. Hence from (16) $\sigma_{U_R}^2$ must lie between $\dfrac{2(n-1)}{n^2(ns-n)}$ and 0. Comparing the maximum possible value, with the value $\dfrac{2(n-1)}{n^2(ns-n+2)}$ of $\sigma_{U_N}^2$ in equation (7), it appears that, if $n(s-1)$ is not too small, σ_{U_R} will never be much greater than σ_{U_N} and hence the chance of rejecting will never much exceed the specified ϵ. We are therefore not likely to err much on the side of overestimating the significance of observed treatment differences. On the other hand if there is too much discrepancy between the variances within the different blocks, σ_{U_R} may be considerably less than σ_{U_N} and the test may seriously underestimate significance. The question must now be asked: how much discrepancy in block variations will be serious?

Our procedure will be to approximate to the U_R-distribution by means of a Type I Pearson-Curve with limits at 0 and 1, i.e. by a curve

$$p(U_R) = \text{const.} \times U_R^{m_1-1}(1-U_R)^{m_2-1}. \qquad \text{......(17)}$$

m_1 and m_2 will be chosen so that the first two moments of this curve agree with the true moments of U_R given by equations (12) and (14). Of course, the distribution of U from randomization must in fact be discontinuous and although U must lie between 0 and 1, these extreme values will never in general be attained. However, although (17) may for these reasons only provide an approximate graduation to the U_R distribution, we may certainly expect it to be better than the normal theory curve of (5), which is of the same form but does not have the correct standard deviation. Certainly for the normal theory to be satisfactory we may demand good correspondence between (5) and (17).

4. *A Particular Example (Randomized Blocks)*. In the following example there are $n=8$ blocks and $s=4$ treatments. This is the case for which T. Eden and F. Yates performed the sampling experiment referred to earlier. f_0 is 21, f_1 is 3 and equation (5) gives

$$p(U_N) = \text{const.} \times U_N^{\frac{3}{2}-1}(1-U_N)^{\frac{21}{2}-1}.$$

From the *Tables of the Incomplete Beta-function** or by transformation from Fisher's z-tables, we find

$$P(U_N > \cdot 305) = \cdot 05; \quad P(U_N > \cdot 410) = \cdot 01.$$

That is to say, ·305 and ·410 are the values of U_0 corresponding to normal theory significance levels of $\epsilon = \cdot 05$ and $\epsilon = \cdot 01$ respectively. Now the m_1 and m_2 of the general type I curve, (17), are connected with its first two moments by the relations

$$m_1 = \frac{\mu_1'(\mu_1' - \mu_2')}{(\mu_2' - \mu_1'^2)}; \quad m_2 = \frac{(1-\mu_1')(\mu_1' - \mu_2')}{(\mu_2' - \mu_1'^2)}. \qquad \text{......(18)}$$

Also in the present case the randomization moments from equations (12) and (14) are

$$\mu_1' = \tfrac{1}{8}; \quad \mu_2' = \frac{5 - 2A}{192}.$$

Hence

$$m_1 = \frac{(19 + 2A)}{16(1-A)}; \quad m_2 = \frac{7(19 + 2A)}{16(1-A)}.$$

For every A in the possible range from $\frac{1}{8}$ to 1 there will be a different approximation (17) to the distribution of U from randomization. For each A, then, we can derive, from the Incomplete Beta-function tables, an approximation to $P(U_R > U_0)$ —the true chance of the first kind of error. This has been done and the results are plotted in Fig. 2. It will first be noticed that the risk of the first kind of error never exceeds by much the value ϵ at which we attempt to fix it. The maximum value occurs at $A = \cdot 125$, where for $\epsilon = \cdot 05$ the risk is ·056 and for $\epsilon = \cdot 01$ it is ·013. The risk decreases as A increases, until at $A = \cdot 192$ it is actually ϵ It decreases further from ϵ to 0 as A increases from ·192 to 1. For values of A within this range the test will tend to underestimate significance.

The data upon which T. Eden and F. Yates performed their sampling experiment were derived from measurements of heights of barley. The eight values of $\sum_j u_{ij}^2$ $(i = 1, 2, \ldots, s)$ for these data are 7628, 15,702, 22,669, 59,732, 3666, 90,593, 26,297 and 8672. By (15) these give $A = \cdot 242$. $\sigma^2_{U_R}$ is ·0079 against the normal theory value ·0084 of $\sigma^2_{U_N}$. From Fig. 2 the risks of the first kind of error are ·046 and ·0085 instead of ·05 and ·01. There is a slight tendency to underestimate significance, but for practical purposes this is negligible.

5. *Further Examples* (*Randomized Blocks*). Whether the test will usually be unbiased depends on the values of A which we are likely to meet in practice. An examination of uniformity trial data in different fields would therefore be of value. In the following I have considered four examples.

(I) A trial with mangolds by A. Mercer and W. Hall published in *Journ. Agric. Sci.* IV (1911), p. 107.

(II) A trial with wheat by A. Mercer and W. Hall in the same paper.

* *Tables of the Incomplete Beta-function*, edited by Karl Pearson, *Biometrika* Office, University College, London.

(III) A trial with oats on the field "Za Baranem" published in the Polish journal *Roczniki Nauk Rolniczych* (1917) and in *Landw. Versuchstationen*, XC (1917), pp. 225–40 (authors, M. Gorski and M. Stefaniow).

(IV) An experiment giving nitrogen content in barley carried out by St Barbacki and published in *Mémoires de l'Institut National Polonais d'Économie Rurale à Pulawy*, XIV, Nr. 213 (1933), pp. 106–57.

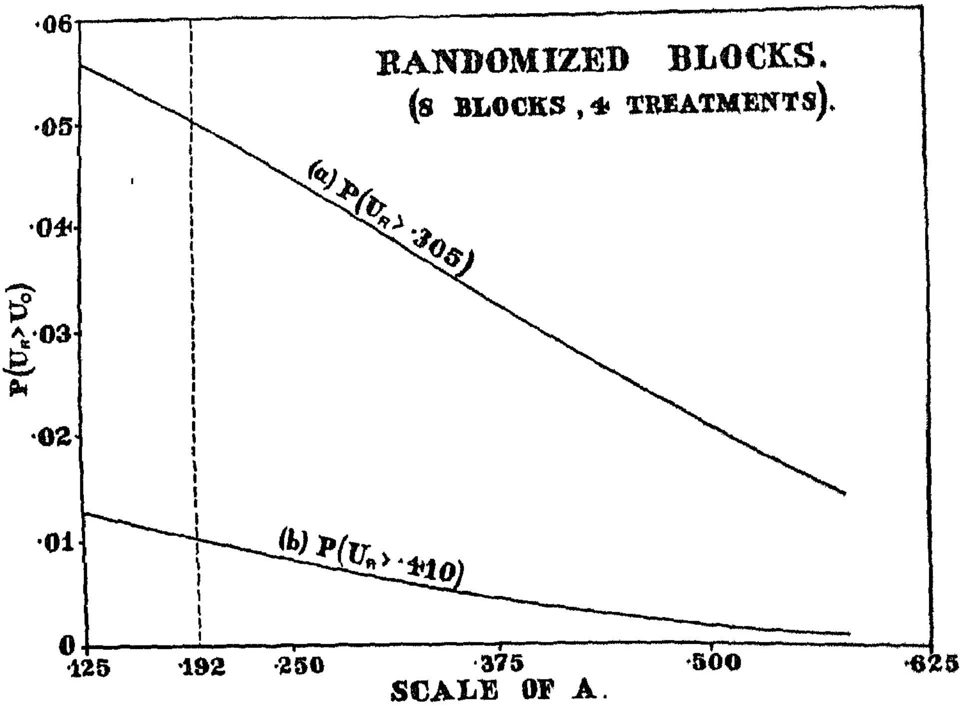


Fig. 2. True probability of the first kind of error in Randomized Block experiment with 8 blocks and 4 treatments. (*a*) Normal theory $\epsilon = \cdot 05$. (*b*) Normal theory $\epsilon = \cdot 01$.

In each case the data as published were grouped up until the plots were of such size and position as might be used in a Randomized Block experiment. For this amalgamated data the necessary information for comparing the U_N and U_R distributions is given in Table II.* It will be noted that for the first three examples there is exceedingly good agreement. For the fourth, the true risks of the first kind of error are ·041 and ·007 instead of ·05 and ·01, and the test tends to underestimate significance.

Whereas these practical trials show no serious bias in the test, it must not be inferred that this will always be the case. Theoretically it has been shown that the test will underestimate significance if the block variances are too discrepant.

* n and s are in each case different from 8 and 4, so that comparison should not be made with Fig. 2.

In the cases where the discrepancies are sufficiently large to matter, the experimenter will probably notice something peculiar in his data and make allowances accordingly. Sometimes, however, there may be doubt and an investigation on the above lines may be useful. It should be noted that the actual work involved is not great. The total "within block" variation is computed in any case for the analysis of variance, and it involves little extra trouble to calculate the separate "within block" variances, from which A is obtained. In this connection a table may be useful showing for different n and s the range of values, A, for which the bias in the test is negligible.

TABLE II

The comparison of U_N and U_R distributions for certain uniformity trial data

Example:	I	II	III	IV
$n \times s$	10×5	12×3	6×5	4×4
$\sigma^2_{U_N}$	·00426	·00588	·01068	·02679
A	·1312	·1847	·2023	·4258
$\sigma^2_{U_R}$	·00434	·00566	·01108	·02392
$p\,(U_R > U_0)$ for $\epsilon = \cdot 05$	·050	·048	·053	·041
$p\,(U_R > U_0)$ for $\epsilon = \cdot 01$	·010	·009	·011	·007

It is of interest to note that the procedure described above, when applied to the example of section 21 of R. A. Fisher's *Design of Experiments*, gives results consistent with his. This example relates to an actual experiment in which the question is asked whether one kind of seed is better than another. A positive difference of means was observed and from the t test there was a chance ·02491 of getting a result as great or greater without there really being a differentiation in seed. By considering the 2^{15} results, which could be obtained in this experiment by randomization on the "null" hypothesis, Fisher obtained, without any approximation, a significance level of ·02628 for the observed difference. By means of the Type I approximation to the U_R distribution I found that, corresponding to the normal theory $\epsilon = \cdot 05$, the chance $p\,(U_R > U_0)$ for this data was ·053. As I was considering the chances of obtaining as large an absolute deviation as the one observed, the result agrees with Fisher's as far as the third decimal place.

6. *Latin Squares* (*Normal Theory*). In Latin Square experiments the field is divided into s rows ($i = 1, 2, \ldots, s$) and s columns ($j = 1, 2, \ldots, s$), making s^2 plots. The treatments tested ($k = 1, 2, \ldots, s$) are arranged on the field so that each falls once and once only into every row and every column. Upon the yields of the experiment the analysis of Table III is performed.

The test, whether there are significant differences between the treatment means, involves the calculation of $z = \frac{1}{2}\log_e(v_1/v_0)$ and reference to tables based on normal theory. This theory proceeds from the assumption that, if the treat-

ments are equivalent, the yield of the plot in the ith row and jth column can be written

$$x = A + R_i + C_j + \eta, \qquad \ldots\ldots(19)$$

where A, R_i $(i = 1, 2, \ldots, s)$ and C_j $(j = 1, 2, \ldots, s)$ are constants, and the η's are normally and independently distributed about zero with the same standard

TABLE III

Analysis of Variance for Latin Square

Source	Degrees of Freedom	Sum of Squares	Mean Square
Between Treatments	$(s-1)$	S_1	v_1
Between Rows	$(s-1)$	S_2	v_2
Between Columns	$(s-1)$	S_3	v_3
Residual	$(s-1)(s-2)$	S_0	v_0
Total	(s^2-1)		

deviation. As in the case of Randomized Blocks, we shall consider $U = (S_1)/(S_0 + S_1)$, which is now related to z by the equation

$$U = \{1 + (s-2)\, e^{-2z}\}^{-1}. \qquad \ldots\ldots(20)$$

On the assumption (19), we have

$$U_N = \frac{\chi_1^2 \sigma^2}{\chi_0^2 \sigma^2 + \chi_1^2 \sigma^2},$$

where χ_0^2 and χ_1^2 are independently distributed as χ^2 with degrees of freedom $f_0 = (s-1)(s-2)$ and $f_1 = (s-1)$ respectively. The distribution of U_N* is therefore

$$p(U_N) = \text{const.} \times U_N^{(f_1/2)-1} (1 - U_N)^{(f_0/2)-1}, \qquad \ldots\ldots(21)$$

and the moments are

$$_N\mu_1' = \frac{1}{(s-1)}, \qquad \ldots\ldots(22)$$

$$_N\mu_2' = \frac{(s+1)}{(s-1)(s^2 - 2s + 3)}, \qquad \ldots\ldots(23)$$

$$\sigma_{U_N}^2 = \frac{2(s-2)}{(s-1)^2 (s^2 - 2s + 3)}. \qquad \ldots\ldots(24)$$

7. *Latin Squares (Randomization Theory).* It is now necessary to consider in more detail the arrangement of the experiment, and to formulate precisely the hypothesis which it is meant to test. Let the yield which the kth treatment is capable of giving on the plot (i, j) be $x_{ij(k)}$. Then we shall suppose that the hypothesis under test is that *every plot* would give the same yield, however treated, i.e. that

$$x_{ij(k)} = x_{ij} \qquad (k = 1, 2, \ldots, s). \qquad \ldots\ldots(25)$$

* The suffix N is used, as before, to denote the distribution of U on normal theory.

We are concerned only with the probability of the first kind of error, and therefore the whole of the following analysis supposes (25) satisfied. As has been stated, the experiment is arranged so that each treatment falls once into every row and every column. For example, suppose the yields x_{ij} form on the field the 4×4 square of Fig. 3 (a), and the arrangement of the treatments on the field follows the plan of Fig. 3 (b). Then the yields reclassified by treatment and row will be as in Fig. 3 (c). The treatment means are 20·25, 24·75, 23·75 and 23, and from them may be calculated the treatment sum of squares, S_1 of Table III. Any other arrangement of treatments, differing from Fig. 3 (b), but satisfying the Latin Square conditions, will lead to a different reclassification of the yields and a different S_1. Randomization enters into the experiment in making the decision as to which particular Latin Square arrangement is to be applied. A fundamental set of possible squares

		Columns (j) 1	2	3	4
Rows (i)	1	30	22	18	28
	2	34	13	21	25
	3	38	12	18	23
	4	33	19	16	17

Fig. 3 (a). Example of yields x_{ij} on a 4×4 field.

		Columns (j) 1	2	3	4
Rows (i)	1	4	2	1	3
	2	1	3	2	4
	3	3	1	4	2
	4	2	4	3	1

Fig. 3 (b). Example of Latin Square arrangement of treatments ($k = 1, 2, 3, 4$).

		Treatments (k) 1	2	3	4
Rows (i)	1	18	22	28	30
	2	34	21	13	25
	3	12	23	38	18
	4	17	33	16	19
Mean		$20\frac{1}{4}$	$24\frac{3}{4}$	$23\frac{3}{4}$	23

Fig. 3 (c). Reclassification of yields obtained on applying Fig. 3 (b) to Fig. 3 (a).

(defined below) is decided upon, and from it one particular square is chosen at random for the experiment. In order to judge, therefore, the significance of the value of the criterion U obtained from an experiment, it is necessary to know something of the distribution of values of U which would be generated if every element of the fundamental set of Latin Squares were applied to the field under essentially the same conditions. As before, attention will here be confined to the first two moments of this distribution, and comparisons will be made with the normal theory moments of (22), (23) and (24).

First the fundamental set of squares must be defined. For s small, this set can be taken to consist of all the different Latin Squares that are possible. Methods of choosing one square at random from this total set have been given by R. A. Fisher and F. Yates for $s \leqslant 6$.* The necessary enumeration of squares, which would make these methods available for $s > 7$, has not yet been performed. Instead, for s

* See F. Yates, "The Formation of Latin Squares for Use in Field Experiments", *Emp. J. exp. Agric.* I (1933), pp. 235–44 and R. A. Fisher and F. Yates, "The 6×6 Latin Squares", *Proc. Camb. phil. Soc.* XXX (1934), pp. 492–507.

from 7 to 12, Yates has simply given a single example of a Latin Square, and suggests that the fundamental set of squares, to be used in randomization, should consist of all squares that can be obtained from this single square, by permutations of rows, columns and treatments. Such a set of squares has been called a *transformation* set.

The mean and second moments of U for squares of a given transformation set will first be considered. It will be found that, for a given s, the mean U for different transformation sets is the same, but this is not the case for the second moment. Knowing the second moment for each transformation set it is not difficult, in the cases when $s \not> 6$, to deduce the second moment over a fundamental set consisting of all squares that are possible. We shall first consider moments over single transformation sets.

It is convenient to write

$$\begin{aligned} x_{ij} &= x_{..} + (x_{i.} - x_{..}) + (x_{.j} - x_{..}) + (x_{ij} - x_{i.} - x_{.j} + x_{..})^* \\ &= A' + R'_i + C'_j + u_{ij} \quad \text{(say)}, \end{aligned} \qquad \ldots\ldots(26)$$

where it will be noted that

$$\left.\begin{aligned} &\sum_i u_{ij} = 0 \quad (j = 1, 2, \ldots, s); \qquad &\sum_j u_{ij} = 0 \quad (i = 1, 2, \ldots, s); \\ &\sum_i R'_i = 0; &\sum_j C'_j = 0. \end{aligned}\right\} \qquad \ldots\ldots(27)$$

Also if the yield of the kth treatment in the ith row is denoted by $x_{i(k)}$, we can write

$$x_{i(k)} = x_{..} + (x_{i.} - x_{..}) + (x_{.j} - x_{..}) + (x_{i(k)} - x_{i.} - x_{.j} + x_{..}),$$

where j is the column into which the kth treatment falls in the ith row. This means that we can write

$$x_{i(k)} = A' + R'_i + C'_j + y_{i(k)} \quad \text{(say)}. \qquad \ldots\ldots(28)$$

It will be seen that only variation in the quantities $y_{i(k)}$ need be considered in the following analysis. The possible sets of values $y_{i(k)}$, which can be obtained by applying the transformation set of squares to the field, are those which can be obtained by applying the squares to the residuals u_{ij}.

The numerator of U is

$$S_1 = s \sum_k \{x_{.(k)} - x_{..}\}^2 = s \sum_k y^2_{.(k)}. \qquad \ldots\ldots(29)$$

The denominator is

$$\begin{aligned} (S_0 + S_1) &= \sum_i \sum_j (x_{ij} - x_{..})^2 - \sum_i s\,(x_{i.} - x_{..})^2 - \sum_j s\,(x_{.j} - x_{..})^2 \\ &= \sum_i \sum_j (x_{ij} - x_{i.} - x_{.j} + x_{..})^2 = \sum_i \sum_j u^2_{ij}, \end{aligned} \qquad \ldots\ldots(30)$$

i.e. is the same no matter what the Latin Square arrangement is. The moments of U depend therefore only on the moments of S_1, and these in turn depend only on the possible sets of y's. From (29) it is seen that S_1 is symmetrical with respect to

* Dots indicate that means are being taken over all the values of the letter replaced by the dot.

treatments. All Latin Squares of the transformation set, which can be derived from one another simply by a permutation of treatments, will therefore give rise to the same value of S_1 and therefore of U. Hence the moments of U over the transformation set are the same as the moments over a set, which can be derived from an original square by permutation of rows and columns only. We shall denote this set by Ω and expectations over Ω by E. Then from (29)

$$\begin{aligned} E(sS_1) &= E[\sum_k \{\sum_i y_{i(k)}\}^2] \\ &= E[\sum_k \sum_i y_{i(k)}^2 + \sum_k \sum_{i \neq m} y_{i(k)} y_{m(k)}] \\ &= \sum_i \sum_j u_{ij}^2 + s \sum_{i \neq m} \left\{ \frac{\sum_{j \neq j'} u_{ij} u_{mj'}}{s(s-1)} \right\},* \end{aligned}$$

since, for any treatment, the column j occupied in row i and the column j' in row m can be with equal likelihood any pair of values, except $j = j'$. Using the relations (27) we obtain

$$\begin{aligned} E(sS_1) &= \sum_i \sum_j u_{ij}^2 + s \sum_{i \neq m} \left\{ \frac{-\sum_j u_{ij} u_{mj}}{s(s-1)} \right\} \\ &= \sum_i \sum_j u_{ij}^2 + \frac{1}{(s-1)} \sum_j \sum_i u_{ij}^2 = \frac{s(\sum_i \sum_j u_{ij}^2)}{(s-1)}. \end{aligned}$$

Hence
$$E(U) = E\left(\frac{S_1}{S_0 + S_1}\right) = \frac{1}{(s-1)}, \qquad \text{......(31)}$$

agreeing with the normal theory value of (22). For the second moment we have

$$\begin{aligned} E(s^2 S_1^2) &= E[\sum_k \{\sum_i y_{i(k)}\}^2]^2 \\ &= E[\sum_k \sum_{k'} \sum_i \sum_m \sum_p \sum_q y_{i(k)} y_{m(k)} y_{p(k')} y_{q(k')}]. \qquad \text{......(32)} \end{aligned}$$

The method used to evaluate this expectation in the case of Randomized Blocks can no longer be applied, since the y's in different rows are not independent. The difficulty is greatest for terms in which $k \neq k'$, e.g. the term $y_{i(1)} y_{m(1)} y_{p(2)} y_{q(2)}$. To obtain the expectation of this, it is convenient to divide the set Ω into a number of sub-sets and first find the expectation for each sub-set. We shall put into one sub-set all the squares of Ω in which treatment 1 is allocated to the same plots. Such a sub-set will be termed $\omega(j_1, j_2, \ldots, j_s)$, where j_i is the column into which treatment 1 falls in the ith row. For example, the square of Fig. 4 (*a*) is a member of $\omega(2, 1, 3, 5, 4)$, and all the members of this set will be obtained by permuting rows and columns in such a way that the 1's are not moved. This permutation may be done by interchanging rows in any way, and then making the necessary column permutation to bring the 1's back to their original positions. If, for

* For summation convention see footnote, p. 26.

example, rows 1 and 4 are interchanged, then necessarily columns 2 and 5 must be interchanged. For different sets ω, the column permutation to be taken in conjunction with any particular row permutation is, of course, different.

4	(1)	3	5	2
(1)	3	2	4	5
3	5	(1)	2	4
5	2	4	3	(1)
2	4	5	(1)	3

Fig. 4 (a). Example of 5×5 square belonging to ω (2, 1, 3, 5, 4).

Row M					
1	(1)	3	2	4	5
2	4	(1)	3	5	2
3	3	5	(1)	2	4
4	2	4	5	(1)	3
5	5	2	4	3	(1)

Fig. 4 (b). Square of same set Ω as Fig. 4 (a), but belonging to ω (1, 2, 3, 4, 5). This particular square may be chosen as the fundamental square of ω (1, 2, 3, 4, 5).

Instead of considering immediately expectations over the general sub-set $\omega\,(j_1, j_2, \ldots, j_s)$, it is simpler to start with the particular set $\omega\,(1, 2, \ldots, s)$, for which treatment 1 lies down the principal diagonal. Expectations over this set will be denoted by E'. Then

$$E'\,[\sum_i\sum_m\sum_p\sum_q y_{i(1)}\,y_{m(1)}\,y_{p(1)}\,y_{q(1)}] = \sum_i\sum_m\sum_p\sum_q u_{ii}\,u_{mm}\,u_{pp}\,u_{qq}$$

$$=(\sum_i u_{ii})^4, \qquad \ldots\ldots(33)$$

and

$$E'\,[\sum_i\sum_m\sum_p\sum_q y_{i(1)}\,y_{m(1)}\,y_{p(2)}\,y_{q(2)}] = (\sum_i\sum_m u_{ii}\,u_{mm})\,E'\,[\sum_p\sum_q y_{p(2)}\,y_{q(2)}]$$

$$=(\sum_i u_{ii})^2\,E'\,[\sum_p\sum_q y_{p(2)}\,y_{q(2)}]. \qquad \ldots\ldots(34)$$

Our first problem, therefore, is to evaluate terms of the form $E'[y_{p(2)}\,y_{q(2)}]$. For, having these, we can deduce (34); then by analogy we can obtain the expectations of (33) and (34) over the more general sub-set $\omega\,(j_1, j_2, \ldots, j_s)$; finally we can combine all the sub-sets* to obtain the expectations of the same quantities over Ω. $E(s^2 S_1^2)$ will then follow from (32).

To fix ideas we shall take one member of $\omega\,(1, 2, \ldots, s)$ and term it the fundamental square of the set. The rows in this square will be numbered $M = 1, 2, \ldots, s$. (Fig. 4 (b) shows this done for squares belonging to the same set Ω as the square of Fig. 4 (a).) The manner in which all the squares of the set can be obtained from the fundamental square is this: the rows can be permuted in any way: the columns must then be permuted in exactly the same way, in order that the 1's should come back on to the principal diagonal. For short, this type of permutation of rows and columns will be termed a symmetrical permutation.

* There will be the same number of different squares in each sub-set $\omega\,(j_1, j_2, \ldots, j_s)$.

First take $q=p$. Then making all symmetrical permutations, treatment 2 in row p will fall with equal frequency into all columns except the pth. Therefore

$$E'\,[y^2_{p(2)}]=(\sum_j u^2_{pj}-u^2_{pp})/(s-1). \qquad \ldots\ldots(35)$$

Next consider $q \neq p$. Treatment 1 will occupy in the rows p and q the positions u_{pp} and u_{qq}. There are now (s^2-3s+3) possible pairs of values for $y_{p(2)}$ and $y_{q(2)}$, but all of these do not occur equally frequently in the set $\omega\,(1, 2, \ldots, s)$. The possible kinds of pairs are illustrated in the following diagram, the bracket denoting the plot on which treatment 2 falls.

$$\begin{array}{llcl}
\text{Row } p \;\ldots\; u_{pp} \;\ldots\; (u_{pq}) \;\ldots\;\ldots\;\ldots & & & \ldots\;\ldots\;\ldots\; u_{pp} \;\ldots\; (u_{pq}) \;\ldots\;\ldots\;\ldots \\
\text{Row } q \;\ldots\; (u_{qp}) \;\ldots\; u_{qq} \;\ldots\;\ldots\;\ldots & & & \ldots\; (u_{qj}) \;\ldots\;\ldots\;\ldots\; u_{qq} \;\ldots\;\ldots\;\ldots \\
\qquad\qquad \text{(i)} & & & \qquad\qquad \text{(ii)} \\
\text{Row } p \;\ldots\; u_{pp} \;\ldots\;\ldots\;\ldots\; (u_{pj}) \;\ldots & & & \ldots\;\ldots\;\ldots\; u_{pp} \;\ldots\;\ldots\;\ldots\; (u_{pj}) \;\ldots \\
\text{Row } q \;\ldots\; (u_{qp}) \;\ldots\; u_{qq} \;\ldots\;\ldots\;\ldots & & & \ldots\; (u_{qj'}) \;\ldots\;\ldots\;\ldots\; u_{qq} \;\ldots\;\ldots\;\ldots \\
\qquad\qquad \text{(iii)} & & & \qquad\qquad \text{(iv)}
\end{array}$$

We may have

$$\begin{array}{lllll}
& \text{(i)} \; y_{p(2)}=u_{pq} & \text{and} & y_{q(2)}=u_{qp}, & \\
& \text{(ii)} \; y_{p(2)}=u_{pq} & \text{and} & y_{q(2)}=u_{qj} & (j \neq p \text{ or } q), \\
& \text{(iii)} \; y_{p(2)}=u_{pj} & \text{and} & y_{q(2)}=u_{qp} & (j \neq p \text{ or } q), \\
\text{or} & \text{(iv)} \; y_{p(2)}=u_{pj} & \text{and} & y_{q(2)}=u_{qj'} & (j \text{ and } j' \neq p \text{ or } q,\; j \neq j').
\end{array}$$

There is 1 pair of kind (i), $(s-2)$ of kind (ii), $(s-2)$ of kind (iii), and $(s-2)(s-3)$ of kind (iv),* making (s^2-3s+3) possible values for $y_{p(2)}\,y_{q(2)}$. To find the relative frequency of these we must now consider certain properties of what we have termed the fundamental square of $\omega\,(1, 2, \ldots, s)$.

Take a particular row M of this square. In this row treatment 1 falls into column M. There will now be some row (R say) for which treatment 2 falls into column M.

	Column M		Column R	
Row M	1		?	
Row R	2		1	

* I am throughout taking s to be greater than 3.

In this row treatment 1 falls into column R. There are now two possibilities: either (*a*) treatment 2 in row M falls into column R, or (*b*) it does not. If it does, we shall say that treatments 1 and 2 satisfy, for row M, the reversal property P. If not, we shall say that they satisfy the property $\bar{P}$. Further, taking all the rows M $(=1, 2, \ldots, s)$, we shall denote by n_{12} the number of them in which the property P is satisfied for treatments 1 and 2. More generally by $n_{kk'}$ will be denoted the number of rows in which treatments k and k' satisfy the reversal property. (For example, in the square of Fig. 4 (*b*), for the treatments 1 and 2, the property P holds for $M=2$ and 5, the property $\bar{P}$ for $M=1$, 3 and 4. Hence n_{12} is 2.)

The relevance of the above considerations is seen when we come to find the proportion of squares of ω $(1, 2, \ldots, s)$ which give $y_{p(2)} y_{q(2)} = u_{pq} u_{qp}$. This is the situation pictured in the first figure of the diagram above. For any square of the set ω $(1, 2, \ldots, s)$ which produces this situation, the four elements falling on the plots u_{pp}, u_{pq}, u_{qp} and u_{qq} satisfy, for treatments 1 and 2, the reversal property. These four elements must therefore correspond to four elements satisfying the reversal property somewhere in the fundamental square of ω $(1, 2, \ldots, s)$, for permutation of rows and columns will not destroy the property. The chance that $y_{p(2)} y_{q(2)} = u_{pq} u_{qp}$, therefore, depends on the number of times the property holds in the fundamental square. Now we have seen that the squares of ω $(1, 2, \ldots, s)$ are obtained by any permutation of rows followed by the same permutation of columns. There are $s(s-1)$ ways in which two rows of the fundamental square can be chosen to fall on the rows p and q of the field. It is seen that the number of these pairs of rows in which treatments 1 and 2 satisfy the reversal property P is n_{12}. Hence the chance of $y_{p(2)} y_{q(2)} = u_{pq} u_{qp}$ in the set ω $(1, 2, \ldots, s)$ is $n_{12}/s(s-1)$.

Let us denote by p_1, p_2, p_3 and p_4 the chances of getting in ω $(1, 2, \ldots, s)$ the individual pairs of values $y_{p(2)}$ and $y_{q(2)}$ referred to above as of kinds (i), (ii), (iii) and (iv). Then we have

$$\left.\begin{aligned} &p_1 + (s-2)p_2 = 1/(s-1); \quad p_2 = p_3; \\ &p_1 + (s-2)p_2 + (s-2)p_3 + (s-2)(s-3)p_4 = 1. \end{aligned}\right\} \qquad \ldots\ldots(36)$$

The first of these relations follows from the fact that in ω $(1, 2, \ldots, s)$ the chance that $y_{p(2)} = u_{pq}$ is $1/(s-1)$. The second follows from considerations of symmetry and the third from the fact that the chances of all $(s^2 - 3s + 3)$ pairs must add up to unity. Using the value of p_1, which we have already evaluated, we obtain from (36)

$$\left.\begin{aligned} &p_1 = n_{12}/s(s-1); \quad p_2 = p_3 = (s - n_{12})/s(s-1)(s-2); \\ &p_4 = \{n_{12} + s(s-3)\}/s(s-1)(s-2)(s-3). \end{aligned}\right\} \qquad \ldots\ldots(37)$$

Now

$$\begin{aligned} E'[y_{p(2)} y_{q(2)}] = p_1 u_{pq} u_{qp} &+ p_2 {\sum_j}' u_{pq} u_{qj} \\ &+ p_3 {\sum_j}' u_{pj} u_{qp} \\ &+ p_4 {\sum_j}' u_{pj} {\sum_{j'}}'' u_{qj'}, \end{aligned}$$

where $\sum_j{}'$ denotes summation of j over all values 1, 2, ..., s, excluding p and q, and $\sum_{j'}{}''$ denotes summation of j' over all values 1, 2, ..., s, excluding p, q and j. Performing these summations, remembering that $\sum_j u_{ij}=0$, and substituting the values of p_1, p_2, p_3 and p_4 from (37), we obtain finally

$$E'[y_{p(2)}y_{q(2)}]=\frac{1}{(s-1)(s-2)}[u_{pq}u_{qq}+u_{pp}u_{qp}-u_{pq}u_{qp}+u_{pp}u_{qq}-\sum_j u_{pj}u_{qj}]$$
$$+\frac{n_{12}}{s(s-1)(s-2)(s-3)}[(s-1)u_{pq}u_{qq}+(s-1)u_{pp}u_{qp}$$
$$+(s^2-3s+1)u_{pq}u_{qp}+u_{pp}u_{qq}-\sum_j u_{pj}u_{qj}]. \qquad \text{......(38)}$$

We have now obtained the expectations of $y^2_{p(2)}$ and $y_{p(2)}y_{q(2)}$ $(q\neq p)$ over the symmetrical permutation set which keeps treatment 1 down the major diagonal. Of the two, the latter depends, not only on the possible yields of the plots in rows p and q, but also on the structure of the squares in Ω. It does not matter which particular square of Ω is chosen to evaluate n_{12}, for the number of rows, in which treatments 1 and 2 satisfy the property P, is quite independent of any permutation of rows and columns (e.g. for any of the squares of Fig. 4, n_{12} is equal to 2).

Next, $E'[\sum_p\sum_q y_{p(2)}y_{q(2)}]$ will be considered, i.e. $\sum_p E'[y^2_{p(2)}]+\sum_{p\neq q}E'[y_{p(2)}y_{q(2)}]$. This will follow from the necessary summations of (35) and (38). These summations are simplified by the following relations:

$$\left.\begin{aligned}\sum_{p\neq q}u_{pp}u_{qq}&=(\sum_p u_{pp})^2-\sum_p u^2_{pp};\\ \sum_{p\neq q}u_{pp}u_{qp}&=\sum_{p\neq q}u_{pq}u_{qq}=-\sum_p u^2_{pp};\\ \sum_{p\neq q}u_{pq}u_{qp}&=\sum_p\sum_q u_{pq}u_{qp}-\sum_p u^2_{pp};\\ \sum_{p\neq q}\sum_j u_{pj}u_{qj}&=-\sum_j\sum_p u^2_{pj}.\end{aligned}\right\} \qquad \text{......(39)}$$

Using these relations, we get finally

$$E'[\sum_p\sum_q y_{p(2)}y_{q(2)}]=\frac{1}{(s-1)(s-2)}[(\sum_p u_{pp})^2-s\sum_p u^2_{pp}-\sum_p\sum_q u_{pq}u_{qp}$$
$$+(s-1)\sum_p\sum_j u^2_{pj}]+\frac{n_{12}}{s(s-1)(s-2)(s-3)}[(\sum_p u_{pp})^2-s(s-1)(\sum_p u^2_{pp})$$
$$+(s^2-3s+1)(\sum_p\sum_q u_{pq}u_{qp})+(\sum_p\sum_j u^2_{pj})]. \qquad \text{......(40)}$$

Substituting (40) into (34), we obtain

$$E'[\sum_i\sum_m\sum_p\sum_q y_{i(1)}y_{m(1)}y_{p(2)}y_{q(2)}]=\frac{1}{(s-1)(s-2)}[X-sY-Z+(s-1)(\sum_i\sum_j u^2_{ij})W]$$
$$+\frac{n_{12}}{s(s-1)(s-2)(s-3)}[X-s(s-1)Y+(s^2-3s+1)Z+(\sum_i\sum_j u^2_{ij})W], \qquad \text{......(41)}$$

where
$$\left.\begin{aligned} X&=(\sum_i u_{ii})^4, \quad Y=(\sum_i u_{ii})^2(\sum_i u_{ii}^2),\\ Z&=(\sum_i u_{ii})^2(\sum_i\sum_m u_{im}u_{mi}), \quad W=(\sum_i u_{ii})^2.\end{aligned}\right\} \qquad \text{......(42)}$$

The expectations of (33) and (41) have been evaluated for the set having treatment 1 down the principal diagonal, i.e. the set for which treatment 1 in row i occupies the column i. For the more general set $\omega(j_1, j_2, \ldots, j_s)$, in which treatment 1 in row i falls on the plot (i, j_i) the expectations will be of exactly the same form, with X, Y, Z and W defined by the more general expressions:

$$\left.\begin{aligned} X&=(\sum_i u_{ij_i})^4, \quad Y=(\sum_i u_{ij_i})^2(\sum_i u_{ij_i}^2),\\ Z&=(\sum_i u_{ij_i})^2(\sum_i\sum_m u_{ij_m}u_{mj_i}), \quad W=(\sum_i u_{ij_i})^2.\end{aligned}\right\} \qquad \text{......(43)}$$

Further, the same expectations for the total set Ω will be obtained by replacing X, Y, Z and W by $E[X]$, $E[Y]$, $E[Z]$ and $E[W]$—their expectations over all the set of permutations $(j_1, j_2, \ldots, j_s)$. Thus

$$E[\sum_i\sum_m\sum_p\sum_q y_{i(1)}y_{m(1)}y_{p(1)}y_{q(1)}]=E[X] \qquad \text{......(44)}$$

and

$$\begin{aligned} E[\sum_i\sum_m\sum_p\sum_q y_{i(1)}y_{m(1)}y_{p(2)}y_{q(2)}]&=\frac{1}{(s-1)(s-2)}\{E[X]-sE[Y]\\ &-E[Z]+(s-1)(\sum_i\sum_j u_{ij}^2)E[W]\}+\frac{n_{12}}{s(s-1)(s-2)(s-3)}\{E[X]-s(s-1)E[Y]\\ &+(s^2-3s+1)E[Z]+(\sum_i\sum_j u_{ij}^2)E[W]\}. \end{aligned} \qquad \text{......(45)}$$

Hence

$$\begin{aligned} E[s^2S_1^2]&=E[\sum_k\sum_{k'}\sum_i\sum_m\sum_p\sum_q y_{i(k)}y_{m(k)}y_{p(k')}y_{q(k')}]\\ &=\frac{s}{(s-2)}\{(s-1)E[X]-sE[Y]-E[Z]+(s-1)(\sum_i\sum_j u_{ij}^2)E[W]\}\\ &+\frac{(\sum_{k\neq k'} n_{kk'})}{s(s-1)(s-2)(s-3)}\{E[X]-s(s-1)E[Y]+(s^2-3s+1)E[Z]\\ &+(\sum_i\sum_j u_{ij}^2)E[W]\}. \end{aligned} \qquad \text{......(46)}$$

The derivation of these expectations of X, Y, Z and W is given in an Appendix to this paper. The results only are presented here. They involve four symmetrical functions of the plot yields, viz.

$$\left.\begin{aligned} D&=\sum_i\sum_j u_{ij}^4, \quad F=(\sum_i\sum_j u_{ij}^2)^2,\\ G&=\{\sum_i(\sum_j u_{ij}^2)^2+\sum_j(\sum_i u_{ij}^2)^2\}, \quad H=\sum_i\sum_m(\sum_j u_{ij}u_{mj})^2.\end{aligned}\right\} \qquad \text{......(47)}$$

With this notation,

$$\left.\begin{aligned}
E[X] &= \frac{1}{s(s-1)(s-2)(s-3)}[s^2(s+1)D+3(s^2-3s+1)F \\
&\qquad\qquad\qquad -3s(s-1)G+6H]; \\
E[Y] &= \frac{1}{s(s-1)(s-2)}[s^2D+(s-1)F-sG]; \\
E[Z] &= \frac{1}{s(s-1)(s-2)(s-3)}[s^2(s+1)D+(s^2-3s+3)F \\
&\qquad\qquad\qquad -3s(s-1)G+2(s^2-3s+3)H]; \\
E[W] &= \frac{\sqrt{F}}{(s-1)}.
\end{aligned}\right\} \quad \ldots\ldots(48)$$

Substituting (48) into (46),

$$E[s^2S_1^2)] = \frac{1}{(s-1)(s-2)^2(s-3)}[2s^2(s-1)D+(s^4-4s^3+2s^2+6s-6)F$$

$$-2s(s^2-3s+3)G-2(s^2-6s+6)H]+\frac{(\sum_{k\neq k'} n_{kk'})}{\{s(s-1)(s-2)(s-3)\}^2}[2s^2(s-1)^2D$$

$$+2(2s^2-6s+3)F-2s(s-1)(s^2-3s+3)G+2(s^4-6s^3+13s^2-12s+6)H]. \quad \ldots\ldots(49)^*$$

The expectation of U^2 then follows from

$$E[U^2]=E\left[\frac{S_1^2}{(S_0+S_1)^2}\right]=\frac{E[S_1^2]}{(\sum_i\sum_j u_{ij}^2)^2}=\frac{E[s^2S_1^2]}{s^2F}. \quad \ldots\ldots(50)$$

In the case of Randomized Blocks the expectation of U^2 depended only on the size of the experiment and on a single function, A, of the plot yields. From algebraic considerations it was possible to show that there was an upper limit to the probability of the first kind of error when the z-test was applied. For the Latin Square the situation is much more complicated. $E[U^2]$ depends on three functions of the plot yields (viz. D/F, G/F and H/F), and also on the function, $(\sum_{k\neq k'} n_{kk'})$, of the structure of a typical square of Ω. No attempt is made here to make any definite statement, which will be independent of the values of these functions, about the probability of the first kind of error. It is possible, however, without much difficulty, to make a direct trial of the applicability of the theoretical z-distribution in any particular instance, by calculating out from the plot yields the quantities D, F, G and H. This has been done in the following section for the data of uniformity trials, and for some hypothetical data in which there are certain systematic fertility gradients.

8. *The* 4×4, 5×5 *and* 6×6 *Squares*. In applying (49) to particular cases, I shall make use of the methods of choosing squares summarized conveniently by F. Yates in the *Empire Journal of Experimental Agriculture*.† The squares

* For confirmation of this result see Appendix B. † *Loc. cit.*

derivable from a single $s \times s$ Latin square by permutation of rows, columns and treatments are called a transformation set. All the $(s!)^3$ permutations do not lead to different squares, but however many different squares there are, each will be repeated the same number of times. A square with 1, 2, 3, ..., s along the first row and 1, 2, 3, ..., s down the first column is called a reduced square. From a reduced square $(s!)(s-1)!$ different squares can be generated by permuting all the rows, except the first, and all the treatments. The number of different squares in a transformation set is equal to the number of different reduced squares in the set multiplied by $(s!)(s-1)!$ There is, in general, more than one transformation set for a given s, but the different transformation sets do not contain the same number of reduced squares. To make a random choice from all the different possible squares of size $s \times s$ (*giving each the same chance of being used*) it is necessary to give each transformation set a chance of being used proportional to the number of reduced squares it contains. In the following we shall consider the distribution of U, firstly over each of the transformation sets, and, secondly, over the whole set of different possible squares of size $s \times s$. Expectations over the set of all possible squares will clearly be weighted means of the expectations over the separate transformation sets—the weights being proportional to the numbers of reduced squares in the sets.

In the cases $s = 4$, 5 and 6, equations (49) and (50) give

$$s=4, \quad \mu_2' = \frac{1}{192F}[50F + 96D - 56G + 4H] + \frac{(\sum_{k \neq k'} n_{kk'})}{9216F}[22F + 288D - 168G + 76H]; \qquad \text{......(51)}$$

$$s=5, \quad \mu_2' = \frac{1}{1800F}[199F + 200D - 130G - 2H] + \frac{(\sum_{k \neq k'} n_{kk'})}{360\,000F}[46F + 800D - 520G + 292H]; \qquad \text{......(52)}$$

$$s=6, \quad \mu_2' = \frac{1}{8640F}[534F + 360D - 252G - 12H] + \frac{(\sum_{k \neq k'} n_{kk'})}{4\,665\,600F}[78F + 1800D - 1260G + 804H]. \qquad \text{......(53)}$$

When $s = 4$ there are two transformation sets, an illustration of each being given in Fig. 5 (*a*). The first set contains 3 different reduced squares and the other only 1. For the first set $(\sum_{k \neq k'} n_{kk'}) = 16$. This is seen in the following way. Consider, say, treatments 2 and 4 in, say, the 3rd row. Treatment 2 falls into the 3rd column. Now see in what row treatment 4 falls into the 3rd column. It is the 2nd row. Now

see if in the 2nd row treatment 2 falls into the same column as treatment 4 does in the 3rd row. It does not. Therefore treatments 2 and 4 do not possess, for row 3, the reversal property, which I have referred to as P. Similarly, this property is seen not to hold for any row, and hence $n_{24}=n_{42}=0$. In the same fashion it is seen that all the other n's are 0, except $n_{12}=n_{21}=4$ and $n_{34}=n_{43}=4$. Thus $(\sum_{k\neq k'} n_{kk'})=16$. In the second set the reversal property holds for all pairs of treatments throughout. All the n's are 4 and thus $(\sum_{k\neq k'} n_{kk'})=48$. The second moment of U about zero for the two sets is therefore obtained by substituting into (51) the values $(\sum_{k\neq k'} n_{kk'})=16$ and 48, respectively. For the second moment over all possible squares we must take $\frac{3}{4}$ of the first result plus $\frac{1}{4}$ of the second, the ratios of the numbers of the reduced squares being as 3 : 1. This is the same thing as substituting into (51) the weighted mean of the $(\sum_{k\neq k'} n_{kk'})$'s, i.e.

$$(3\times 16+1\times 48)/4=24.$$

I

1	2	3	4
2	1	4	3
3	4	2	1
4	3	1	2

II

1	2	3	4
2	1	4	3
3	4	1	2
4	3	2	1

Fig. 5 (a). Illustrations of two 4×4 transformation sets.

I

1	2	3	4	5
2	1	4	5	3
3	5	1	2	4
4	3	5	1	2
5	4	2	3	1

II

1	2	3	4	5
2	3	4	5	1
3	4	5	1	2
4	5	1	2	3
5	1	2	3	4

Fig. 5 (b). Illustrations of two 5×5 transformation sets.

For the 5×5 squares there are again two transformation sets, illustrated in Fig. 5 (b). In the first set there are 50 reduced squares and $(\sum_{k\neq k'} n_{kk'})=16$. In the second set there are 6 reduced squares and $(\sum_{k\neq k'} n_{kk'})=0$. The weighted mean of $(\sum_{k\neq k'} n_{kk'})$ is $(50\times 16+6\times 0)/56$, i.e. $14\frac{2}{7}$. The second moment of U about zero, over the two sets and over all possible squares, is obtained by substituting respectively these three numbers for $(\sum_{k\neq k'} n_{kk'})$ in (52).

For the 6×6 squares there are 22 transformation sets. Yates illustrates only 17 of these, since the other 5 can be obtained by rotating 5 of the 17 through a right angle. For our purpose also it is unnecessary to distinguish between two sets of squares, one of which is the other rotated through a right angle. Such rotation does not affect $(\sum_{k\neq k'} n_{kk'})$. The numbers of reduced squares and $(\sum_{k\neq k'} n_{kk'})$ for each of the 17 sets is given in Table IV. Substitution into (53) gives the corresponding μ_2' for U. The least $(\sum_{k\neq k'} n_{kk'})$ is 0, the greatest 108, and the weighted mean $33\frac{9}{147}$.

TABLE IV

Summary of 6 × 6 transformation sets

Yates Index Number	I	II	III	IV	V	VI	VII	VIII	IX
Number of Reduced Squares	2160	1080	1080	1080	1080	540	540	720	360
$(\sum_{k \neq k'} n_{kk'})$	20	20	16	44	28	28	76	60	60
Yates Index Number	X	XI	XII	XIII	XIV	XV	XVI	XVII	
Number of Reduced Squares	180	240	120	60	40	72	36	20	
$(\sum_{k \neq k'} n_{kk'})$	36	36	36	36	0	60	60	108	

9. *Examples (Latin Squares).* In the following section the examples are numbered and described. It may here be noted that, although we may derive $s!(s-1)!$ different squares from a reduced square by complete permutation of treatments and permutation of all rows except the first, all these squares do not give a different U. For if the hypothesis that the treatments are equivalent is true, any permutation of treatments will not affect the between treatment sum of squares and therefore will not affect U. A reduced square therefore gives only $(s-1)!$ different U's. When $s=4$, this means that one transformation set gives $3 \times 3! = 18$ values and the other $1 \times 3! = 6$ values. The complete set of possible values is 24. For $s=4$, therefore, it is not difficult to work out all the possible values and derive therefrom second moments from randomization. This was done in the first example to test the correctness of (49).

Example I (4×4). Artificially constructed set of values u_{ij} of Fig. 6 (*a*).

*Example II** (4×4). Uniformity trial giving nitrogen content in barley (St Barbacki).

*Example III** (5×5). Uniformity trial with oats (Gorski and Stefaniow).

*Example IV** (6×6). Uniformity trial with wheat (Mercer and Hall).

Example V (6×6). An artificial set of yields x_{ij}, given in Fig. 6 (*b*), in which the fertility level runs diagonally across the field.

Example VI (6×6). An artificial set of yields x_{ij} given in Fig. 6 (*c*), in which the yield on any plot is equal to the yield on the plot two columns to the right in the next row.

For each of the above examples the necessary functions of the plot yields were computed and substituted in the appropriate equation (51), (52) or (53). For the 6×6 squares the equation (53) was not evaluated for all the transformation sets but only for the ones giving the extreme values of 0 and 108 to $(\sum_{k \neq k'} n_{kk'})$. The

* Examples II, III, IV are regroupings of the same data used in Examples IV, III and II of section 5 of the paper. The necessary references are given there. The Mercer and Hall wheat yields are given in Table VI.

expectation for the set of all possible squares was also evaluated. The results are given in Table V. The variance, $\sigma^2_{U_N}$, of U from normal theory is of course obtained from equation (24). In the last column but one is entered the ratio of the variance from randomization to the variance from normal theory, i.e. $\sigma^2_{U_R}/\sigma^2_{U_N}$.

(a)

1	0	3	−4
2	−2	0	0
3	1	−2	−2
−6	1	−1	6

(b)

50	41	32	17	65	71
65	50	41	32	17	65
47	65	50	41	32	17
23	47	65	50	41	32
38	23	47	65	50	41
41	38	23	47	65	50

(c)

50	41	32	17	65	77
65	59	50	41	32	17
46	51	65	59	50	41
24	19	46	51	65	59
35	29	24	19	46	51
32	47	35	29	24	19

Fig. 6. (a) Set of residuals u_{ij} of yields on artificial 4×4 field.
(b) and (c) Artificial sets of yields x_{ij} on 6×6 field.

TABLE V

Comparison of $\sigma^2_{U_R}$ and $\sigma^2_{U_N}$ for artificial examples and for data of three uniformity trials

Example	s	$\sum_{k\neq k'} n_{kk'}$	μ_2' from randomization	$\sigma^2_{U_R}$	$\sigma^2_{U_N}$	$\sigma^2_{U_R}/\sigma^2_{U_N}$	$P(U>U_0)$ for $\epsilon=\cdot 05$
I	4	16	·13914	·02803	·04040	·6938	—
		48	·13066	·01955	·04040	·4838	—
		24*	·13702	·02591	·04040	·6413	—
II	4	16	·13218	·02107	·04040	·5215	—
		48	·12417	·01306	·04040	·3233	—
		24*	·13018	·01907	·04040	·4720	—
III	5	16	·07809	·01559	·02083	·7481	—
		0	·07862	·01612	·02083	·7740	—
		$14\frac{2}{7}$*	·07814	·01564	·02083	·7509	·0288
IV	6	0	·048215	·008215	·011852	·6931	—
		108	·049106	·009106	·011852	·7683	—
		$33\frac{9}{147}$*	·048487	·008487	·011852	·7161	·0271
V	6	0	·05355	·01355	·011852	1·1432	—
		108	·05403	·01403	·011852	1·1839	—
		$33\frac{9}{147}$*	·05370	·01370	·011852	1 1556	·0624
VI	6	0	·05139	·01139	·011852	·9609	—
		108	·05415	·01415	·011852	1·1937	—
		$33\frac{9}{147}$*	·05223	·01223	·011852	1·0321	·0528

For each example two different transformation sets are considered and also the set of all possible squares. This latter is indicated by an asterisk in column 3. In the last column is given the probability of the first kind of error when the ϵ of normal theory is ·05.

For the uniformity trials II, III and IV it is seen that there is a very much greater disparity between the variances from the normal and randomization theories than was the case for Randomized Blocks. This seems to indicate that the z-test in the Latin Square is more liable to bias. Admittedly only three examples are considered here, but the data in two of them, at least, are not exceptional. Mercer and Hall's wheat data have often been used in discussions of the present character and I give in Table VI (a) the actual grouped yields which I used in Example IV. I made slight adjustments in these yields to make the sums of rows and columns divisible by 6. This made it simpler to evaluate the residuals $u_{ij}=(x_{ij}-x_{i\cdot}-x_{\cdot j}+x_{\cdot\cdot})$. Some of the residuals turned out to be three-figure numbers, and as this makes the evaluation of D, F, G and H rather heavy I rounded them off to the two-figure numbers given in Table 6 (b).* These adjustments do not, I think, affect materially the ratios D/F, G/F and H/F.

TABLE VI

Mercer and Hall's wheat data

(a)

37·47	36·68	37·61	36·73	35·72	32·96
36·20	35·94	36·82	36·71	33·43	36·51
34·00	35·80	37·23	36·63	34·93	38·42
34·20	37·89	37·40	36·75	32·67	34·71
37·83	38·05	35·56	38 39	31·75	32·16
38·50	35·60	37·63	37·27	33·01	28·68

(b)

0·7	−0·4	0·2	−0·8	1·7	−1·4
−0·3	−0·9	−0·4	−0·5	−0·3	2·4
−2·8	−1·3	−0·2	−0·0	0·9	4·3
−2·0	1·4	0·5	−0·2	−0·8	1·1
1·6	1·5	−1·3	1·5	−1·7	−1·6
2·8	−0·3	1·2	0·9	0·2	−4·8

Mercer and Hall give wheat yields in lb. for 500 plots of $\frac{1}{500}$ acre each. By taking the first 18 rows and the first 18 columns of this data and regrouping in 36 bigger square plots of size $\frac{1}{180}$ acre, a 6×6 square with the yields in (a), above, was obtained. In (b) are given the residuals when row and column variation is allowed for. Certain adjustments made in deriving these residuals are referred to in the text.

It is an interesting point that the two systematic arrangements of Examples V and VI give good agreement between normal and randomization theories.

For the 5×5 and 6×6 squares here considered, the size of $(\sum_{k\neq k'} n_{kk'})$ does not seem to matter much. The expectations are much the same for all the transformation sets. For larger squares than $s=6$ the differences between transformation sets will probably become still less important.

In the last column of Table V I have given my approximation to the true probability of the first kind of error, when the rejection level is based on the 5 per cent. point U_0 of normal theory. These levels are obtained by approximating to the U distribution from randomization, by means of a Pearson type I curve (as in the case of Randomized Blocks). I have not done this in the examples with $s=4$, since there are only 24 possible values of U and the approximation by a continuous curve does not seem justified. In the other examples I give the risk of rejecting only when the randomization set consists of all possible squares of the size.

* I did the same thing with Examples II and III. Without these simplifications the numbers would have become uncomfortably large.

Results very little different would be obtained for the individual transformation sets. For the uniformity trials III and IV the probabilities of the first kind of error are ·029 and ·027 respectively, instead of the required ·05. There is in these cases a definite underestimation of significance by the usual z-test. This statement must, however, be qualified by the remarks in the next section.

10. *Summary and Conclusions.* In experiments in which randomization is performed, the actual arrangement of treatments on the field is one chosen at random from a predetermined set of possible arrangements. In the present paper investigation has been made for Randomized Blocks and Latin Square experiments, into the distribution of the statistic z, generated by the application to the observed plot yields of the whole fundamental set of arrangements, assuming as true the "null" hypothesis that the treatments have no differential effect on the plots. It was found convenient to consider, instead of z, a monotonically increasing function U of z, which is equal to the treatment sum of squares divided by the total of the treatment sum of squares and the residual sum of squares.

Comparison of the U distribution from randomization with that from normal theory showed, in both Randomized Blocks and Latin Square, exact agreement of the means, but disagreement in the variances with consequent disagreement in the proportions of the distribution falling beyond certain points. Some uniformity trial data were used in order to see whether, in practice, these disagreements were of sufficient magnitude to be of importance. For Randomized Blocks the cases considered showed close enough agreement between the randomization and normal theory variances of U. In each of three uniformity trials for Latin Square, however, the randomization variance of U was considerably smaller than that of the normal theory. Whether this should be taken as evidence of bias in the usual z-test based on normal theory, depends on the point of view adopted concerning the hypothetical population about which the data of an experiment is supposed to give information. Let us consider two possibilities.

(i) We may make, from the yields of the experiment, a *statistical* inference only about the situation on the particular field of the experiment, e.g. as in the present paper, we may be using our *statistical* method only to test whether all the treatments would have given identical yields on each plot of this particular field. Of course if we come to the conclusion that the treatments can be regarded as equivalent on this field, we probably make the further induction that they can be regarded as equivalent over some wider range of experience. Otherwise the experiment would be useless. However, from the present viewpoint, this further inference is not a statistical one in the usual sense. If the statistical part of the inferences made from an experiment go no farther than the experimental field, then in cases where the variance of U from randomization on the "null" hypothesis is less than that from normal theory, we may say that the usual z-test underestimates

the significance of the observed treatment differences. Where the variance of U from randomization is the larger, there the z-test overestimates significance.

(ii) We may, alternatively, choose to regard the inferences, drawn from the experimental yields to some wider experience, as being *completely* statistical. This means that we shall not only regard the statistic z calculated from the yields as a random sample from the distribution, which is obtained on the "null" hypothesis by applying all the possible arrangements of the fundamental set to the experimental field; we shall in turn regard this randomization distribution of z as a random sample from a set of similar distributions, which might be obtained in other experiments of a similar type. These may be carried out on different fields, under differing weather conditions, and may be subject to different technical errors in the harvesting and weighing of the crop and so on. Clearly with this wider conception, the results derived here for three Latin Square uniformity trials are insufficient to give any general answer to the question of bias arising in the application of the z-test. We should note, however, that randomization ensures the agreement of the mean U with normal theory. The second moment of U for *individual* fields may, as we have seen, differ appreciably from the normal theory value. In a number of experiments, however, these differences may tend to balance out, so that on the average the discrepancy may be negligible and the normal theory test unbiased.

In this connection it is of interest to recall the investigation of O. Tedin,* who considered the application of 5×5 Latin Squares to 91 uniformity trials. He took twelve different arrangements of the 5×5 square. Each arrangement was applied to all the 91 trials, giving for each arrangement 91 values of a criterion, which is practically the same as my U and which he termed a "treatment error coefficient". He came to the conclusion that it was dangerous to apply systematically the same arrangement (at least if it was of either the Diagonal or the Knight's Move pattern) in every experiment and still expect the normal theory z-test to be unbiased. The application of the methods of the present paper to such a set of uniformity trials would, I think, be useful. It would indicate how far the process of randomization does actually eliminate bias, when the z-test is regarded from the second viewpoint mentioned above.

APPENDIX

A. *Derivation of Expectations in Equation* (48). In the following, $\sum'$ will be used to denote summation over all possible sets of values of the row suffixes, excluding terms in which two or more of the row suffixes are the same. Thus, for example, $\sum' u_{ij_i}^2 u_{lj_l} u_{mj_m}$ is a summation over all i, l and m, excluding terms in

* O. Tedin, "The Influence of Systematic Plot Arrangements upon the Estimate of Error in Field Experiments", *J. agric. Sci.* XXI (1931), p. 191.

which two or more of i, l and m are equal. This summation, therefore, has $s(s-1)(s-2)$ terms. With this notation,

$$W = (\sum_i u_{ij_i})^2 = \sum' u_{ij_i}^2 + \sum' u_{ij_i} u_{lj_l};$$

$$Y = (\sum_i u_{ij_i})^2 (\sum_i u_{ij_i}^2) = \sum' u_{ij_i}^4 + \sum' u_{ij_i}^2 u_{mj_m}^2 + 2\sum' u_{ij_i}^3 u_{lj_l} + \sum' u_{ij_i} u_{lj_l} u_{mj_m}^2;$$

$$X = (\sum_i u_{ij_i})^4 = \sum' u_{ij_i}^4 + 4\sum' u_{ij_i}^3 u_{lj_l} + 3\sum' u_{ij_i}^2 u_{lj_l}^2 + 6\sum' u_{ij_i}^2 u_{lj_l} u_{mj_m} + \sum' u_{ij_i} u_{lj_l} u_{mj_m} u_{rj_r};$$

$$Z = (\sum_i u_{ij_i})^2 (\sum_i \sum_l u_{ij_l} u_{lj_i}) = \sum' u_{ij_i}^4 + 2\sum' u_{ij_i}^3 u_{lj_l} + \sum' u_{ij_i}^2 u_{mj_m}^2 + 2\sum' u_{ij_i}^2 u_{ij_r} u_{rj_i} + \sum' u_{ij_i}^2 u_{mj_r} u_{rj_m} + \sum' u_{ij_i}^2 u_{lj_l} u_{mj_m} + 2\sum' u_{ij_i} u_{lj_l} u_{ij_l} u_{lj_i} + 4\sum' u_{ij_i} u_{lj_l} u_{ij_r} u_{rj_i} + \sum' u_{ij_i} u_{lj_l} u_{mj_r} u_{rj_m}. \qquad \ldots\ldots(54)$$

W, X, Y and Z are thus dependent on twelve kinds of term, which are listed in column 1 of Table VII. The expectations of these terms are all derived in the same manner, and, as the algebra is somewhat long, only one example will be given in full here. For instance, consider $E(u_{ij_i}^2 u_{mj_r} u_{rj_m})$. The expectation is being taken over all sets of values $(j_1, j_2, \ldots, j_s)$ which are permutations of the numbers $(1, 2, \ldots, s)$. The term under present consideration involves only the three different rows i, m and r, and we have only to consider what happens when j_i, j_m and j_r take all the values $(1, 2, \ldots, s)$, excluding any two of them being equal. Let us start by taking j_i and j_r fixed. Then j_m can take all values $(1, 2, \ldots, s)$, except j_i and j_r. Hence

$$E[u_{ij_i}^2 u_{mj_r} u_{rj_m}] = E\left[u_{ij_i}^2 u_{mj_r} \frac{(\sum_j u_{rj} - u_{rj_i} - u_{rj_r})}{(s-2)}\right]$$
$$= -\frac{1}{(s-2)}\{E[u_{ij_i}^2 u_{rj_i} u_{mj_r}] + E[u_{ij_i}^2 u_{mj_r} u_{rj_r}]\}.$$

Now consider j_i fixed, so that j_r can take all values $(1, 2, \ldots, s)$ except j_i. Hence

$$E[u_{ij_i}^2 u_{mj_r} u_{rj_m}] = -\frac{1}{s-2}\left\{E\left[u_{ij_i}^2 u_{rj_i} \frac{(\sum_j u_{mj} - u_{mj_i})}{(s-1)}\right] + E\left[u_{ij_i}^2 \frac{(\sum_j u_{mj} u_{rj} - u_{mj_i} u_{rj_i})}{(s-1)}\right]\right\},$$

i.e. $$E[u_{ij_i}^2 u_{mj_r} u_{rj_m}] = \frac{1}{(s-1)(s-2)}\{E[2u_{ij_i}^2 u_{rj_i} u_{mj_i}] - (\sum_j u_{mj} u_{rj}) E[u_{ij_i}^2]\}.$$

Finally, j_i can take all values $(1, 2, \ldots, s)$. Hence

$$E[u_{ij_i}^2 u_{mj_r} u_{rj_m}] = \frac{1}{s(s-1)(s-2)}\{2\sum_j u_{ij}^2 u_{rj} u_{mj} - (\sum_j u_{mj} u_{rj})(\sum_j u_{ij}^2)\}.$$

The same method leads to the other entries in column 2, the essential point of the work being the repeated use of the relation $\sum_j u_{ij} = 0$. Next we have to consider

TABLE VII

(1)	(2) Expectation of (1)	(3) Expectation of (1) summed over Σ'
$u^2_{ij_i}$	$\frac{1}{s}(\sum_j u^2_{ij})$	$\frac{1}{s}\sqrt{F}$
$u_{ij_i}u_{lj_l}$	$-\frac{1}{s(s-1)}(\sum_j u_{ij}u_{lj})$	$\frac{1}{s(s-1)}\sqrt{F}$
$u^4_{ij_i}$	$\frac{1}{s}(\sum_j u^4_{ij})$	$\frac{1}{s}D$
$u^3_{ij_i}u_{lj_l}$	$-\frac{1}{s(s-1)}(\sum_j u^3_{ij}u_{lj})$	$\frac{1}{s(s-1)}D$
$u^2_{ij_i}u^2_{lj_l}$	$\frac{1}{s(s-1)}\{(\sum_j u^2_{ij})(\sum_j u^2_{lj})-(\sum_j u^2_{ij}u^2_{lj})\}$	$\frac{1}{s(s-1)}\{F+D-G\}$
$u^2_{ij_i}u_{lj_l}u_{mj_m}$	$\frac{1}{s(s-1)(s-2)}\{2\sum_j u^2_{ij}u_{mj}u_{lj}-(\sum_j u_{lj}u_{mj})(\sum_j u^2_{ij})\}$	$\frac{1}{s(s-1)(s-2)}\{4D+F-2G\}$
$u^2_{ij_i}u_{ij_r}u_{rj_i}$	$-\frac{1}{s(s-1)}(\sum_j u^3_{ij}u_{rj})$	$\frac{1}{s(s-1)}D$
$u^2_{ij_i}u_{mj_r}u_{rj_m}$	$\frac{1}{s(s-1)(s-2)}\{2\sum_j u^2_{ij}u_{rj}u_{mj}-(\sum_j u_{mj}u_{rj})(\sum_j u^2_{ij})\}$	$\frac{1}{s(s-1)(s-2)}\{4D+F-2G\}$
$u_{ij_i}u_{lj_l}u_{ij_l}u_{lj_i}$	$\frac{1}{s(s-1)}\{(\sum_j u_{lj}u_{ij})^2-(\sum_j u^2_{ij}u^2_{lj})\}$	$\frac{1}{s(s-1)}\{H+D-G\}$
$u_{ij_i}u_{lj_l}u_{ij_r}u_{rj_i}$	$\frac{1}{s(s-1)(s-2)}\{2\sum_j u^2_{ij}u_{rj}u_{lj}-(\sum_j u_{ij}u_{lj})(\sum_j u_{ij}u_{rj})\}$	$\frac{1}{s(s-1)(s-2)}\{4D+H-2G\}$
$u_{ij_i}u_{lj_l}u_{mj_m}u_{rj_r}$	$\frac{1}{s(s-1)(s-2)(s-3)}\{(\sum_j u_{lj}u_{mj})(\sum_j u_{ij}u_{rj})+(\sum_j u_{ij}u_{lj})(\sum_j u_{mj}u_{rj})$ $+(\sum_j u_{ij}u_{mj})(\sum_j u_{lj}u_{rj})-6\sum_j u_{ij}u_{lj}u_{mj}u_{rj}\}$	$\frac{3}{s(s-1)(s-2)(s-3)}\{2H+F+12D-6G\}$
$u_{ij_i}u_{lj_l}u_{mj_r}u_{rj_m}$	$\frac{1}{s(s-1)(s-2)(s-3)}\{(\sum_j u_{lj}u_{mj})(\sum_j u_{ij}u_{rj})+(\sum_j u_{ij}u_{lj})(\sum_j u_{mj}u_{rj})$ $+(\sum u_{ij}u_{mj})(\sum u_{lj}u_{rj})-6\sum_j u_{ij}u_{lj}u_{mj}u_{rj}\}$	$\frac{3}{s(s-1)(s-2)(s-3)}\{2H+F+12D-6G\}$

the expectations of the elements in column 1, when the summation $\sum'$ is applied to them. This involves the evaluation of summations $\sum'$ of the quantities listed in column 1 of Table VIII. Again, only one example will be given here in full, to

TABLE VIII

(1)	(2) Column (1) summed over Σ'
$(\sum_j u^2_{ij})$	$\sqrt{F}$
$(\sum_j u_{ij}u_{lj})$	$-\sqrt{F}$
$(\sum_j u^4_{ij})$	D
$(\sum_j u^3_{ij}u_{lj})$	$-D$
$(\sum_j u^2_{ij})(\sum_j u^2_{lj})$	$F-G'$
$(\sum_j u^2_{ij}u^2_{lj})$	$G''-D$
$(\sum_j u^2_{ij}u_{mj}u_{lj})$	$2D-G''$
$(\sum_j u^2_{ij})(\sum_j u_{lj}u_{mj})$	$2G'-F$
$(\sum_j u_{lj}u_{ij})^2$	$H-G'$
$(\sum_j u_{ij}u_{lj})(\sum_j u_{ij}u_{rj})$	$2G'-H$
$(\sum_j u_{ij}u_{lj})(\sum_j u_{mj}u_{rj})$	$2H+F-6G'$
$(\sum_j u_{ij}u_{lj}u_{mj}u_{rj})$	$3G''-6D$

illustrate the method employed. The essential feature is the repeated use of the relation $\sum_i u_{ij}=0$. For instance, take $\sum'(\sum_j u_{ij}u_{lj})(\sum_j u_{ij}u_{rj})$. First keep i and l fixed and sum r over all values $(1, 2, \ldots, s)$, excluding i and l:

$$\sum'(\sum_j u_{ij}u_{lj})(\sum_j u_{ij}u_{rj}) = \sum'(\sum_j u_{ij}u_{lj})(\sum_j u_{ij}\overline{-u_{ij}-u_{lj}})$$
$$= -\sum'(\sum_j u_{ij}u_{lj})(\sum_j u^2_{ij}) - \sum'(\sum_j u_{ij}u_{lj})^2.$$

Now keep i fixed and sum l over all values $(1, 2, \ldots, s)$, excluding i:

$$\sum'(\sum_j u_{ij}u_{lj})(\sum_j u_{ij}u_{rj}) = -\sum'(\sum_j u^2_{ij})(\sum_j u_{ij}\overline{-u_{ij}}) - \{\sum_i\sum_l(\sum_j u_{ij}u_{lj})^2 - \sum_i(\sum_j u^2_{ij})^2\}$$
$$= 2\sum_i(\sum_j u^2_{ij})^2 - \sum_i\sum_l(\sum_j u_{ij}u_{lj})^2.$$

The results of the other similar summations are given in column 2 of Table VIII, use being made of the following notation:

$$D=\sum_i\sum_j u^4_{ij}, \qquad F=(\sum_i\sum_j u^2_{ij})^2,$$
$$G'=\sum_i(\sum_j u^2_{ij})^2, \qquad G''=\sum_j(\sum_i u^2_{ij})^2, \qquad H=\sum_i\sum_m(\sum_j u_{ij}u_{mj})^2,$$
$$G=G'+G''=\{\sum_i(\sum_j u^2_{ij})^2+\sum_j(\sum_i u^2_{ij})^2\}.$$

From the expressions of columns 2 of Tables VII and VIII, the expectations given in column 3 of Table VII are deduced. These are the expectations of the different kinds of terms involved in W, X, Y and Z, and by substitution into (54) the expectations $E[W]$, $E[X]$, $E[Y]$ and $E[Z]$ are obtained. These are given in equations (48) of the paper.

B. *Confirmation of Equation* (49). The algebraic processes leading to equation (49) are so heavy that one would feel more confident of their correctness if some practical test were made. This can very easily be done when $s=4$. For then there are only 24 possible values of S_1 and these can be calculated directly from the data. This was done in Example I of section 9 and exact agreement with the theory was observed. For $s>4$ the number of possible Latin Squares seemed too large to permit a complete investigation of this kind. It is possible, however, to obtain some general confirmation in the following way.

In the Randomization theory we made no assumptions about x_{ij}. Let us now consider what happens if we apply the reasoning of the theory to a situation where the x's do actually satisfy the equation

$$x_{ij}=A+R_i+C_j+\eta_{ij},$$

the η's being normal independent variates with mean zero and common standard deviation σ. One set of values x_{ij} satisfying these conditions may be termed a *configuration*. There are possible an infinite number of such configurations. We shall denote expectations over all these by E''.

Now consider the set Ω of possible Latin Square arrangements which can be applied to the values x_{ij}. Whatever individual square of Ω is applied to the set x_{ij} it is clear that, in repeated configurations, the resultant values of S_1 will be distributed as $\chi^2\sigma^2$ and therefore $E''[S_1^2]=(s^2-1)\sigma^4$. Hence, if $E[S_1^2]$ denotes the expectation of S_1^2 over all the Latin Squares of Ω applied to the *same* configuration, we must have *a fortiori*

$$E''\{E[S_1^2]\}=(s^2-1)\sigma^4, \qquad \text{i.e.} \quad E''\{E[s^2S_1^2]\}=s^2(s^2-1)\sigma^4.$$

But $E[s^2S_1^2]$ is given by the right-hand side of (49). Hence, if we take E'' {right-hand side of (49)}, we should obtain $s^2(s^2-1)\sigma^4$. Now it can be shown that

$$E''\{D\}=E''\{\sum_i\sum_j(x_{ij}-x_{i\cdot}-x_{\cdot j}+x_{\cdot\cdot})^4\}=3(s-1)^4\sigma^4/s^2,$$

$$E''\{F\}=(s-1)^2(s^2-2s+3)\sigma^4,$$

$$E''\{G\}=2(s-1)^3(s+1)\sigma^4/s, \quad E''\{H\}=(s-1)^2(2s-1)\sigma^4.$$

Substituting these for D, F, G, H in (49), we do in fact get $s^2(s^2-1)\sigma^4$. This provides a check on the accuracy of the formula, although, of course, it does not constitute a proof of its correctness.

SOME ASPECTS OF THE PROBLEM OF RANDOMIZATION

By E. S. PEARSON

1. Introductory

The practical problem of mathematical statistics is to provide a conceptual model which will be of value to the man who needs to draw conclusions from the data of observation. In handling statistical data one of the commonest problems to be faced is that of drawing inferences from a part to the whole, from a sample to the population; such inferences are uncertain inferences, and it follows not only that in such cases the conceptual model must be constructed with the aid of the theory of probability but that its value to the practical man will be to some extent psychological. An historical study of the development of mathematical statistics shows an ever-increasing complexity in the structure of the abstract model and also an evolution of ideas as to how that model is to be of most use in practical application. In this course of evolution it is inevitable that many different suggestions should have been thrown out by mathematical statisticians as to the best way of linking the world of concepts with the world of experience. Ultimately, it is likely that the practical scientist, who may know relatively little mathematics but has to apply the methods of statistics in his research work, will play the decisive part in determining the form in which the theory of probability may be applied most usefully in different situations as a guide to judgment. But in the meantime it is necessary that amid the growing complication of the mathematical background statisticians should attempt to keep clear the simple principles which in their view have the greatest claim for acceptance.

An example of the gradual evolution of ideas is found in the changing attitude with which tests of goodness of fit and tests to determine whether differences are "significant" have been regarded. Perhaps one may say that 20 or 30 years ago the question posed by the statistician in applying such tests was often somewhat as follows:

"If my sample had come (*a*) from the population represented by my fitted curve, or (*b*) from a population whose parameters had the values given by the sample (and these estimates obtained from the sample cannot be very different from the unknown population values), what is the probability that a difference as great or greater than that observed would have occurred?"

It will be seen that the situation posed was to some extent hypothetical, since in fact the population sampled was not represented by the sample values. Nevertheless, the probability measure, P, obtained as an answer to this question

seemed to give the measure of assurance needed to make a decision. In so far as each problem was considered in isolation from other similar problems, the basis for any decision taken was to a large extent psychological.

In recent years we can follow the gradual introduction of a somewhat different conception. Its origin may be traced partly to the application of statistical methods in new fields where decisions had to be taken on evidence supplied by small samples, so that the differences between population values and sample estimates became so large that the hypothetical situation referred to above was seen to be noticeably unreal; and partly to the fact that in agricultural research investigations precisely similar tests were being applied again and again to the same type of experiment. Thus the relation was emphasized between (*a*) the probability measure, P, leading to a decision in an individual experiment, and (*b*) the expected proportion of times that a hypothesis of "no difference" would be wrongly rejected in the routine work of a research station. In terms of the older approach there might be little difference in an isolated problem between the psychological reaction to a P of ·05 and a P of ·02. But where experimental procedures were being repeated continually, the difference between a risk of mistake of 1 in 20 and of 1 in 50 might be of some consequence.*

Emphasis was therefore given in statistical literature to a new idea; that of planning a sampling procedure and the subsequent analysis of the data collected, in such a way as to control at any desired level the risk of making a wrong decision—that risk which can never be entirely eliminated in any form of work involving sampling. This change in attitude is illustrated by the form which many recently constructed probability tables have taken, following R. A. Fisher's suggestion. Instead of providing the statistician with the precise value of a probability measure, P, which he needed when regarding each problem in isolation, these tables are arranged so as to enable him to discover whether his test criterion falls below a certain "probability level", e.g. a 10, 5, 2, or 1 per cent. level. If then, for example, as a usual practice he rejects the hypothesis he is testing when the criterion falls below the 1 per cent. level (but not otherwise), he knows that in the long run of his experience this action will lead to one wrong decision in every hundred, a frequency of error which he may be quite prepared to accept.

This form of introduction of abstract theory into the world of experience has an obvious appeal to the practical man. If you tell him that theory enables him to assess the probability of a certain event in an individual trial or even to assess the frequency with which it would occur under somewhat hypothetical conditions, he may be unconvinced of the value of this theory to him. But if you can illustrate the statistician's objective by two examples of the following type, you are much more likely to convince him of the value of statistical tools.

* This has been brought out very clearly in questions of routine sampling in industry.

Example 1. A frequent problem is one in which, having a sample of n values of a variable x, it is wished to determine limits between which the unknown population mean, ξ, almost certainly lies. Under certain conditions the statistician can here provide a rule for determining from the sample data two limits ξ_a and ξ_b, such that the statement

$$\xi_a \leqslant \xi \leqslant \xi_b$$

may be made with a specified measure of confidence. The details of the procedure advocated depend on the application of the twofold principle that in making such a statement we are concerned, (a) to know the percentage of times it will be correct in long-run application under appropriate conditions, (b) to make the interval in some way as narrow as possible To reduce the risk of error and to reduce the breadth of interval are, beyond a certain point, conflicting objectives and a balance must be struck between them, the statistical method shows how this may be done.

Example 2. Another common problem is one in which two samples are available and it is wished to test the hypothesis that they have been drawn from populations having the same means, $\xi_1 = \xi_2$. Again, the statistician can under certain conditions give a rule of procedure suggesting when the hypothesis should be rejected, and he may base this on another twofold principle: arrange so that in the long run application of the rule, (a) the hypothesis of "no difference" in means will only be rejected when it is true on a small and known percentage of occasions; (b) the hypothesis will be rejected as often as possible when there is a true difference in means, i.e. when $\xi_1 - \xi_2 \neq 0$.

These conceptions have no doubt always been present in the minds of mathematical statisticians but they have only been given precise formulation in recent years. The principle illustrated in Example 1 forms the basis of J. Neyman's work on confidence intervals and the confidence coefficient (1), and although presented in somewhat different form, I think, underlies R. A. Fisher's conception of fiducial probability (2). The principle mentioned in Example 2 forms the basis of J. Neyman and the present writer's work on the testing of statistical hypothesis (3), (4), but in the application of the conception (b) we are at variance with R. A. Fisher. In our view, just as in the simple problem of "interval estimation" mentioned in Example 1, it is necessary to specify the form of population distribution before the interval ξ_a, ξ_b can be calculated, so it is only possible to determine in any precise manner which is the most efficient test of a hypothesis if we can specify the class of alternative hypotheses. Thus, following quite simple principles, we may construct in the conceptual workshop the tests most appropriate in different precisely defined situations. It is then for the practical man to decide which of these situations corresponds most closely to that with which he is faced.

In Fisher's view the experimenter cannot and need not define the alternatives to the hypothesis he is testing. Indeed, Fisher would seem to consider it to be

important to a test of significance that it should be free from the necessity of introducing any elaborate type of background or alternatives which might be true. While I agree that the experimenter cannot specify all conceivable alternatives to the hypothesis tested, I think that a study of the situations met with in practice suggests that he does in fact usually have a fairly clear idea of the alternatives most likely to be true, and that if the mathematical statistician enables him to use this knowledge in picking out the most efficient statistical tool, he will be grateful. If it can be shown that in the situation most likely to exist (e.g. normal variation) one test will detect the falsity of the hypothesis of "no difference" more often than any other test, the appeal in favour of its adoption will surely be very strong.

2. Randomization

I have referred to the idea of arranging a sampling procedure so that conclusions drawn upon application of an appropriate statistical technique will be subject to a known and controlled risk of error. The principle of randomization, whose introduction is largely due to R. A. Fisher, provides a device to aid in the achievement of this objective. Most of the statistical tests used in the more complex sampling problems have been developed on the assumption that the variables are normally distributed, and while it is often clear that considerable departure from normality will not seriously effect their validity, it may be asked how far can tests be constructed which are completely independent of any assumption of normality?

Fisher has given an interesting illustration of such a test based on randomization in section 21 of his book, *The Design of Experiments* (5). The example is suggested by an investigation of Darwin's into the growth-rate of crossed and self-fertilized plants.

In the arrangement of the experiment fifteen seeds resulting from each type of fertilization were used; denote these by A-type and B-type seeds. Fifteen pairs of plots, say p_{si} ($s=1, 2, \dots 15$, $i=1, 2$), were chosen and prepared in such a way that the environmental conditions within each pair were as alike as possible. Following the principle of randomization it would then be necessary to determine at random, and for each pair independently, which site should be occupied by A and which by B-type seed.* After the experiment was completed, the grown plants were measured; suppose the character considered (height at given age) had values of a_s and b_s respectively, for the sth pair of plots. Darwin's problem was to determine whether there was any evidence that the type of fertilization affected the vigour of the plant. Statistically, this can be examined by testing the hypothesis, say H_0, that as far as the character measured on the grown plants is concerned, the two samples of seeds have been drawn from identical populations.

* This process of random assignment was not, of course, actually performed by Darwin.

The character in the grown plant depends on (i) the environmental conditions which may be slightly different between plots p_{s1} and p_{s2}, (ii) some quality inherent in the individual seed. We may now imagine in the conceptual field a continued repetition of the experiment, fifteen seeds of each type being randomly selected, fifteen pairs of plots being prepared and a random assignment of the seeds. If then the method of fertilization is unconnected with subsequent growth, a given quality of seed will be as likely to be associated with an A as a B-type seed, and owing to the randomization will be as likely to be associated with the environmental condition of plot p_{s1} as plot p_{s2}. Hence a difference

$$x_s = a_s - b_s$$

of given numerical magnitude will be as likely to be positive as negative. This will be true independently for all fifteen plots.

It follows that the conceptual population of possible experimental results $x_1, x_2, \ldots, x_{15}$ may be divided into an infinite number of subpopulations each defined by a given set of fifteen values of $|x_s|$, and each containing the 2^{15} elements that will be generated by assigning to these numerical values all possible combinations of positive and negative signs. If the hypothesis, H_0, of no differentiation between the A-type and B-type seed populations is true, each of these 2^{15} elements is equally likely to arise.

To construct a test it is now necessary to find a rule, *applicable to every one of these subpopulations*, which will divide the 2^{15} elements into two classes:

(1) a class I containing a proportion P of the elements,

(2) a class II containing a proportion $1-P$ of the elements.

If then we reject the hypothesis, H_0, of no differentiation when the element represented by the fifteen differences $x_1^0, x_2^0, \ldots, x_{15}^0$, actually observed falls into class I, but not otherwise, we may be sure that the risk of rejecting H_0 when it is true is controlled at a value of P: e.g. if $P = \cdot 05$ we should be using what is ordinarily termed a 5 per cent. significance level. The practical question is, of course, how to determine classes I and II. Clearly they should be so determined that if one type of seed in fact produces larger plants than the other, the element represented by the observed differences $x_1^0, \ldots, x_{15}^0$ would be likely to fall into class I, and thus H_0 would, correctly, be rejected. It is seen at once that some consideration of the alternatives to the hypothesis tested is entering into the construction of the test; it has already entered into the design of the experiment since the care taken to make the environmental conditions associated with the pair of plots p_{s1} and p_{s2} similar, was aimed at increasing the chance of detecting a true difference in seed type if one exists.

Fisher's suggestion is to put into class I the $100P$ per cent. of the 2^{15} elements or a number as near that figure as possible for which the fifteen x's have the largest numerical mean value. Thus, for the data of Darwin's experiment, the

values of $x_1^0, x_2^0, \ldots, x_{15}^0$ were: 49, -67, 8, 16, 6, 23, 28, 41, 14, 29, 56, 24, 75, 60, -48, giving a mean of $314/15 = 20{\cdot}9\dot{3}$.

Taking the subpopulation of $2^{15} = 32{,}768$ elements generated by all possible assignments of positive and negative signs to the fifteen values of $|x_s^0|$, Fisher finds that in 1722, or 5·26 per cent., the numerical value of the mean, $\bar{x}$, is greater than the observed value, $20{\cdot}9\dot{3}$. Consequently, the observed result falls just outside the class I associated with a 5 per cent. significance level and we should probably not be prepared to risk rejecting the hypothesis H_0.

The test proposed by Fisher depends upon a particular definition of the class I. It is important to note that this definition is in no sense unique. For example, we could have put into class I the $100P$ per cent. of the 2^{15} elements for which the geometric mean of the fifteen values $(100 + x_s)$ differed most from 100. I do not suggest that this would be a rational classification, but it is worth while reflecting whether, if we choose to use the arithmetic mean as criterion, we are not being influenced, perhaps unconsciously, by

(*a*) the knowledge that if variation is normal, a criterion based on the observed mean difference in samples will be most efficient in detecting a real population difference in seed types;

(*b*) the belief that the characters measured, a_s and b_s, are likely to be approximately normally distributed.

If this is the case, it would seem that the usefulness of the test is in fact dependent on the form of the alternative hypotheses.

Another illustration of the application of this principle of randomization has been recently given by Fisher elsewhere(6). He supposes we have available measures of the stature of a random sample of, say, n Frenchmen and n Englishmen, and wish to test the hypothesis that the mean height of the sampled populations of Frenchmen and Englishmen are identical. Let the observations be written as follows:

$$\begin{cases} \text{Frenchmen } x_1, x_2, \ldots, x_n, \text{ Mean } \bar{x}; \\ \text{Englishmen } y_1, y_2, \ldots, y_n, \text{ Mean } \bar{y}. \end{cases}$$

If the $2n$ observations were written on cards and shuffled without regard to nationality, it would be possible to divide them into a group A and a group B, each containing n cards, in $(2n)!/(n!)^2$ ways. For each way of division we shall have a mean $\bar{a}$ for group A and a mean $\bar{b}$ for group B, giving a difference

$$d = \bar{a} - \bar{b}.$$

Just as in the last example, divide these $(2n)!/(n!)^2$ possible differences into

(1) class I containing the $P\,(2n)!/(n!)^2$ (or a number as near below this as possible) giving largest values of $|d|$,

(2) class II containing the remaining cases.

Suppose P is chosen to be ·05. Then if the difference $\bar{x} - \bar{y}$ for the observed

French-English subdivision falls into class I, we may reject the hypothesis tested, knowing that the risk we run of rejecting it when it is in fact true is ·05 or less.

This ingenious suggestion of Fisher's leads to the following result: if we adopt the rule wherever a problem of this type arises in our statistical experience, we shall have precise control of the risk of wrong rejection no matter what was the type of variation in the populations sampled.

Of course the procedure needed to determine whether the observed sample falls into class I or class II is very lengthy, unless the samples are very small. I am concerned, however, not with this point, but with the question of whether there is something fundamental about the form of the test suggested, so that it can be used as a standard against which to compare other more expeditious tests, such as Student's. It seems to me that Fisher is overstating the claim of an extremely ingenious device when he writes ((6), p. 59): "Actually, the statistician does not carry out this very simple and very tedious process, but his conclusions have no justification beyond the fact that they agree with those which could have been arrived at by this elementary method" The following example should at any rate help to bring out some points which appear to need careful consideration.

The figures given below represent two samples of seven observations from two populations; they form Experiment I of Table I.

Sample 1. 45, 21, 69, 82, 79, 93, 34. Mean $=\bar{x}_1=60\cdot43$. Midpoint between extreme values $=m_1=57$.

Sample 2. 120, 122, 107, 127, 124, 41, 37. Mean $=\bar{x}_2=96\cdot86$. Midpoint between extreme values $=m_2=82$.

After pooling these fourteen numbers, they can be redivided into two groups A and B, of seven each, in $(14!)/(7!)^2=3432$ ways. We may now ask in how many of these ways:

(1) the difference in means of the two groups has an equal or greater negative value than the observed

$$\bar{x}_1-\bar{x}_2=60\cdot43-96\cdot86=-36\cdot43;$$

(2) the difference in midpoints has an equal or greater negative value than the observed

$$m_1-m_2=57-82=-25?$$

After a rather troublesome investigation into the possible arrangements I find the answer to question (1) is 126 out of 3432 or 3·67 per cent., and to question (2) is 45 out of 3432 or 1·31 per cent. It may be said therefore that random assignments of the fourteen numbers into two groups of seven would give (1) as large or a larger numerical value than that observed to the difference in *means* on 7·3 per cent. of occasions, and (2) as large or a larger numerical value to the difference in *midpoints* on 2·6 per cent. of occasions. It follows that in

applying this form of test to the midpoints, we should be more likely to suspect a difference in populations sampled than in applying the test to the means.

Now of course it is quite possible that in individual cases an inferior test may detect a real difference when a better test does not. I give below therefore four further pairs of random samples from the same two populations, as well as the

TABLE I

Experimental Sampling Data

Experiment	I		II		III		IV		V	
Sample	1	2	1	2	1	2	1	2	1	2
	45	120	29	50	14	60	47	60	67	47
	21	122	41	125	70	104	4	90	18	71
	69	107	27	112	32	81	49	84	41	43
	82	127	5	86	79	41	49	100	41	115
	79	124	27	40	87	69	23	93	65	66
	93	41	58	98	25	40	52	32	8	124
	34	37	92	50	2	48	67	98	52	56
Mean	60·43	96·86	39·86	80·14	44·14	63·29	41·57	79·57	41·71	74·57
Midpoint	57·0	82·0	48·5	82·5	44·5	72·0	35·5	66·0	37·5	83·5

TABLE II

Number of pairs of samples, under randomization, having negative values for $\bar{x}_1-\bar{x}_2$ and m_1-m_2 as great as or greater than the observed pairs

	Mean			Midpoint		
	Greater difference	Equal difference*	Total	Greater difference	Equal difference*	Total
Experiment I	121	5	126	40	5	45
II	56	1	57	44	10	54
III	Over 250	$\geqslant 3$	>253	100	41	141
IV	17	3	20	17	14	31
V	82	2	84	28	25	53

* Including the observed difference itself.

results of applying the two tests. It will be seen that in only one case out of the five does the mean supply stronger evidence of difference than the midpoint. Both these tests are equally valid in the sense that, using either, we can control the error of rejecting the hypothesis that the populations are the same when it is in fact true. In the case taken the population means were at 49·5 and 79·5 respectively and their two standard deviations were the same (=28·86).

Yet as far as the very limited experimental evidence goes, the midpoint test has been the more effective in detecting the presence of the real difference of 30 units in population means. The reason for this is explained at once when we know that the population distributions were rectangular, e.g.

for population 1 any value of x between 00 and 99 was equally likely to occur;
for population 2 any value of x between 30 and 129 was equally likely to occur.*

Since the standard error of the midpoint in samples of n from a rectangular population of standard deviation σ is

$$\sigma_m = \sigma\sqrt{\frac{6}{(n+1)(n+2)}},$$

which for $n=7$ is $\cdot 289\sigma$; while for the mean

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}},$$

which for $n=7$ is $\cdot 378\sigma$; we should expect on theoretical grounds that the difference in sample midpoints, rather than in sample means, would be more efficient in detecting real differences in population means. Such a property would certainly appeal to the practical experimenter, were not both tests for other reasons too lengthy to carry out as a common practice.

Now of course in practice it is extremely unlikely that we should deal with variables whose probability distribution is rectangular, but I have introduced these examples because it seems to me to suggest that in problems of this kind it is impossible to make a rational choice between alternative tests unless we introduce some information beyond that contained in the sample data, i.e. some information as to the kind of alternatives with which we are likely to be faced.

If the variation is approximately normal and the standard deviations in the two populations are the same, the advantages of Student's t-test can be expressed in simple terms which appeal to the practical statistician. Its use gives control of the risk of rejecting the hypothesis of "no difference" when it is true, and at the same time makes more probable than does any other test the detection of a real difference in means.† It is certainly possible to claim that these reasons justify its use rather than the relation it bears to the test of Fisher's which I have outlined. It is true that when variation departs from the normal the t-test will not give quite accurate control of the risk of wrong rejection of H_0 (although the error will usually be small), while the test based on randomization will continue to do so. It is in this that the value of the randomization test lies; but as I have pointed out, in so far as this latter test is applied to *means*, it cannot be regarded as unique, and for wide departures from normality it could probably be improved on by use of other central estimates.

* Tippett's Random Sampling Numbers were used; *Tracts for Computers*, No. XV.

† For discussion of this conception see (3) and (4).

3. Randomization applied to the Latin Square

The conceptual model which lies behind the design of the Latin Square experiment leads to the following expression for the yield on the plot in the ith row and jth column receiving the kth treatment:

$$y_{ij(k)} = A + R_i + C_j + T_k + \eta_{ijk}.$$

Here, for a given experiment with an $s \times s$ Latin Square, A, R_i, C_j and T_k ($i = 1, \ldots, s, j = 1, \ldots, s$; $k = 1, \ldots, s$) may be regarded as constants and the η's as normally and independently distributed about zero with standard deviation, σ_η. The hypothesis, H_0, which it is generally wished to test is that $T_k = 0$ ($k = 1, \ldots, s$), i.e. that there are no treatment differences.

It has always been recognized, however, that the additive row and column contributions, R_i and C_j, given in this equation cannot provide sufficient elasticity to fit all forms of fertility gradient found in practice. Consequently there is bound to be some correlation among the η's from neighbouring plots, and further the η's may not be normally distributed. In a single experiment it is of course quite impossible to decide whether the s^2 values of η can be reasonably regarded as independent normal deviates. Two lines of procedure seem therefore to have been followed.

In the first place emphasis has been laid on the importance of randomization; in assigning the s treatments to their plots, the particular Latin Square pattern used is chosen at random from the very many possible patterns, say N_s in number. The infinite population of results which can be conceived as obtainable from the experiment, if H_0 is true, may then be divided into an infinite set of subpopulations, each containing a finite number of elements, N_s. Each subpopulation is defined by a set of s^2 yields, y_{ij}, and an element corresponds to a partition of these yields into s treatment groups in accordance with a particular one of the N_s Latin Square patterns. The observed result following from the Latin Square pattern chosen for the experiment represents a single one of these elements.

If now, as far as yield is concerned, the s treatments are identical, it will follow that each of these N_s elements is equally likely to occur owing to the random choice of patterns, even if the η's are not normal or independent. Consequently, as in the previous illustrations, it is only necessary to find a rule, applicable to all sets of s^2 yields, which will enable us to separate from the N_s elements a suitable class I containing a proportion P of them. If this can be done, and the hypothesis of no treatment differences is rejected when the experiment performed gives a result falling into this class, we shall run a risk equal to P of rejecting the hypothesis H_0 when it is true.

Exactly as in the simpler examples, many ways might be found of classifying N_s partitions of the yields, y_{ij}; the choice between them may be influenced by expediency or by the efficiency of the resulting test in detecting the presence of

real treatment differences when they exist. From both points of view it seems reasonable to employ the usual z-criterion, although as soon as we must depart from the original model of the equation above, the fundamental association between sums of squares and normal variation is blurred. Accepting this criterion, class I will consist of the PN_s partitions leading to the largest values of z.

In his paper (7) published on pp. 21–52 above, B. L. Welch has suggested a method of determining approximately the lower limit of z bounding this class and he finds that, if for example $P = {\cdot}05$ or ${\cdot}01$, this limit does not necessarily correspond exactly to the 5 and 1 per cent. significance levels found from the usual tables of the z probability integral. Where it falls will in fact depend upon the particular set of s^2 yields, y_{ij}. Thus, in one example taken, as few as 2·8 per cent. and in another as few as 2·9 per cent. of the N_s partitions obtained by randomization of yields from a uniformity trial gave values of z above the normal theory 5 per cent. level. This line of approach suggests, therefore, that if we are to obtain a correct probability level for z from the classification of the N_s partitions, it might be necessary to apply a somewhat lengthy procedure to each set of s^2 yields obtained from an experiment.

The second method of attack is one which, while recognizing that the η's may not be exactly independent or normal, asks how far an analysis of uniformity trial data (for which the T_k in the equation are zero) suggests that the distribution of z differs at all seriously from the normal theory form. In this case only a single z is obtained from each experiment, and we are concerned with the distribution of z resulting from experiments which have actually been carried out, rather than that generated hypothetically under randomization when all possible N_s partitions are obtained from the s^2 yields of a single experiment. The investigation carried out by O. Tedin (8) showed that for certain types of Latin Square pattern the distribution of z found in 91 uniformity trials was definitely biased, but for other patterns selected at random this bias was not evident.

It should be noted that even if the assumptions underlying the Latin Square equation were perfectly satisfied, there can be little doubt as a result of B. L. Welch's work that certain sets of plot yields will occur in practice from time to time which, under randomization, will lead to distributions of z differing from normal theory. Some of these distributions, however, would be biased in one way, some in another, so that when they are all combined together the resulting z-distribution should approach that of normal theory. From each randomization set the experimenter is concerned in fact with only one value of z, and this has been selected at random if he has chosen his Latin Square pattern randomly; consequently from the point of view of his long-run experience, the appropriate probability distribution for him to use would appear to be that of normal theory.*

* Possibly we have here another instance of the difference referred to above between regarding a test as giving essentially a rule to be applied and justified by long-run experience, rather than a probability measure associated with an isolated experiment.

On the other hand if the equation fails to represent the situation commonly met with in the field in such a way that there is a general bias in one direction, the resulting under (or over) estimate of significance could be avoided by the lengthy process of referring the observed z in each case to its appropriate randomization distribution.

To throw further light on these points it would certainly seem to be of interest to extend Welch's investigation by applying his results to further uniformity trial data.

The conception of randomization illustrated in the examples given above is both exceedingly suggestive and often practically useful, but perhaps it should be described as a valuable device rather than a fundamental principle. Its adoption, when it can be followed by the calculation necessary to determine what I have described as the class I elements, ensures accuracy in the determination of the probability level of a test criterion, but without the aid of some further principle it cannot help us to decide which of a number of alternative tests to choose. It seems hardly possible to build the methods of statistics into a consistent whole without facing squarely the why of that choice.

REFERENCES

(1) J. Neyman (1934). *J. R. statist. Soc.* xcvii, Appendix, p. 589, and (1937) *Phil. Trans. Roy. Soc.* (at press).
(2) R. A. Fisher (1930). *Proc. Camb. phil. Soc.* xxvi, 528.
(3) J. Neyman and E. S. Pearson (1933). *Philos. Trans.* A, ccxxxi, 289.
(4) —— —— (1936). *Statistical Research Memoirs* (Dept. of Statistics, University College, London), i, 1.
(5) R. A. Fisher (1935). *The Design of Experiments.* Oliver and Boyd.
(6) —— (1936). *J. R. anthrop. Inst.* lxvi, 57.
(7) B. L. Welch (1937). *Biometrika*, xxix, 21.
(8) O. Tedin (1931). *J. agric. Sci.* xxi, 191.

THE DISTRIBUTION OF THE RATIO OF COVARIANCE ESTIMATES IN TWO SAMPLES DRAWN FROM NORMAL BIVARIATE POPULATIONS

By H. O. HIRSCHFELD,

Harper-Adams Agricultural College, Newport, Shropshire

It is well known that the analysis of variance of a single variable necessitates a test of significance, for which Fisher's z-test is the appropriate solution. However, when problems in more than one variable arise, we must consider in addition to the separate variances the question of correlation and covariation. For every kind of analysis the subdivision of the sum of products of the deviations from the respective means into its different components may be performed in exactly the same way as the subdivision of the sum of squares, and what is generally known as an "analysis of variance and covariance" can be worked out easily.

There are three types of problems for which covariance estimates may be used in a test:*

(i) The question whether there is a difference between the regression coefficients of the two normal populations, from each of which we assume one of the samples has been drawn. This test is related to the theory of residual variance in an analysis of variance and covariance.

(ii) The question whether there is a difference between the correlations in the two above populations.

(iii) The question whether there is any difference between corresponding second order parameters.*

In this paper we are mainly interested in question (ii), though the practical example to be considered later is an example for both question (i) and question (ii). We do not deal with question (iii).

A difference between two correlations is most conveniently tested by Fisher's z-transformation of the estimated correlation coefficient. One great advantage of this test is that it is entirely independent of the values of the population variances, i.e. it is valid when nothing whatever is known about the values of the variances.

There are, however, cases in which the estimated variances allow the assumption that corresponding population variances are equal. Such information, however, which may be derived from the variance estimates is purposely ignored

* See, however, the paper by E. S. Pearson and S. S. Wilks (6), where several other problems are dealt with.

when testing correlations estimated from such samples by Fisher's z-transformation.

Assuming, as we may, that corresponding variances of the populations, from which such samples have been drawn, are equal, a difference between their covariances has exactly the same meaning as a difference between their correlations.

This is the type of problem for which a test for a difference between covariances has been developed in this paper with the help of the distribution of the ratio of covariance estimates.*

We shall not enter into a detailed discussion of the appropriateness of this test. It has been said that the distribution of the ratio of covariance estimates is extremely complicated and that it depends on the population parameters. We frankly admit that it essentially depends on the value of the population correlation ρ, and that there is less scope for its common application than for the test of the z-transformation. But it may be regarded as the object of this paper to make a test of this kind at all possible by showing that the distribution of the ratio of covariance estimates may be developed in such a way that its dependence on ρ is of fairly simple character, and by demonstrating how it may be applied to practical examples.

Thus our problem, well defined by the underlying "null-hypothesis", may be stated as follows:

Let x_i, y_i $(i=1, 2, \ldots, n')$ and X_j, Y_j $(j=1, 2, \ldots, N')$ be two samples both drawn from the same normal bivariate population† and let

$$v=(n'-1)^{-1}\sum_{i=1}^{n'}(x_i-\bar{x})(y_i-\bar{y}),$$

$$V=(N'-1)^{-1}\sum_{j=1}^{N'}(X_j-\bar{X})(Y_j-\bar{Y})$$

be the respective estimates of the covariance. We then ask for the chance P (say) of drawing two samples of the above sizes from the population such that the ratio

$$c=v/V$$

is greater (or less) than a certain value c_0 (say). Knowing the value of P for every value of c_0, we then judge the significance of the observed ratio by substituting it for c_0 and comparing the corresponding value of P with the standard levels of ·05 and ·01.

In this paper we shall find the solution of this problem on the assumption that both n' and N', the numbers of items in the samples, are odd numbers. This restriction, unimportant for practical applications, simplifies the mathematical

* In terms of the paper by E. S. Pearson and S. S. Wilks (6) the situation may be characterized by saying that among the set Ω of all pairs of normal populations π_1, π_2 with parameters ξ_{x_1}, ξ_{y_1}, σ_{x_1}, σ_{y_1}, ρ_1 and ξ_{x_2}, ξ_{y_2}, σ_{x_2}, σ_{y_2}, ρ_2 respectively, for which the relations $\sigma_{x_1}=\sigma_{x_2}$, $\sigma_{y_1}=\sigma_{y_2}$ are fulfilled, the hypothesis is tested that in addition $\rho_1=\rho_2$.

† Throughout the paper normal populations are called identical if they have equal variances and correlations.

expression for P considerably, and thus makes it possible to use the latter for a test in practical examples, as will be shown in this paper.

We shall state here the mathematical expression for the chance P, giving the proof in a mathematical appendix. This expression is given in terms of complete and incomplete B-functions, for which tables have been provided by K. Pearson (1). Using his notation we define

$$\mathrm{B}(p,q)=\int_0^1 x^{p-1}(1-x)^{q-1}\,dx=(p-1)!\,(q-1)!/(p+q-1)!\,,^*$$

$$\mathrm{B}_X(p,q)=\int_0^X x^{p-1}(1-x)^{q-1}\,dx,$$

$$I_X(p,q)=\mathrm{B}_X(p,q)/\mathrm{B}(p,q).$$

We are now ready to write down:

(*a*) The chance $P^+(\rho)$ that two independent samples, viz.

$$x_i, y_i\ (i=1,2,\ldots,\ n'=2k+3),$$
$$X_j, Y_j\ (j=1,2,\ldots,N'=2K+3),$$

which have been drawn from the same normal population, with correlation coefficient ρ, yield estimates of covariance whose ratio is greater than a certain positive value c_0. We have

$$P^+(\rho)=4^{-k-K-2}k^{-1}K^{-1}\sum_{q=1}^{k+1}\sum_{p=1}^{K+1} I_X(p,q)\frac{2^{p+q}}{\mathrm{B}(k,k+2-q)\,\mathrm{B}(K,K+2-p)}$$
$$\times(1-\rho^2)^{k+K+2}\{(1-\rho)^{-p-q}+(1+\rho)^{-p-q}\},\quad \ldots\ldots(1)$$

where $$X=(K+1)/\{(K+1)+c_0(k+1)\}.$$

(*b*) The chance $P^-(\rho)$ that the above samples yield estimates of covariance whose ratio is smaller than a certain negative value c_0.

$$P^-(\rho)=4^{-k-K-2}k^{-1}K^{-1}\sum_{q=1}^{k+1}\sum_{p=1}^{K+1} 2^{p+q}/[\mathrm{B}(k,k+2-q)\times\mathrm{B}(K,K+2-p)]$$
$$\times(1-\rho^2)^{k+K+2}$$
$$\times\{(1-\rho)^{-p}(1+\rho)^{-q}I_{X_1}(p,q)+(1+\rho)^{-p}(1-\rho)^{-q}I_{X_2}(p,q)\},\quad \ldots\ldots(2)$$

where $$X_1=(K+1)(1-\rho)/\{(K+1)(1-\rho)+c_0(k+1)(1+\rho)\},$$
$$X_2=(K+1)(1+\rho)/\{(K+1)(1+\rho)+c_0(k+1)(1-\rho)\}.$$

The expressions for $P(\rho)$ are thus *finite* weighted sums of incomplete B-functions, the weights being complete B-functions and simple polynomials in ρ^2. For $\rho=0$ the chances (1) and (2) are identical, i.e. the distribution is symmetrical in c_0. As ρ^2 tends to 1 (i.e. as ρ tends to $+1$ or to -1) equation (1) approaches the simplified form

$$P^+(\pm 1)=I_X(k+1,K+1),\quad \ldots\ldots(1')$$

* For integral values of p and q.

which is the representation of Fisher's z-test in terms of Pearson's incomplete B-functions. Thus, whenever it is known that there is a very high correlation between the variates x and y, covariance estimates might approximately be tested like mean squares by Fisher's z-test.

On the other hand, since expression (2) approaches 0 as $\rho^2 \to 1$, negative values of c_0 are less and less likely to occur as ρ^2 increases and there will be but few problems with reasonably correlated variates, in which (2) has to be used for a test.

From formulae (1) and (2) it is clear that the chance P essentially depends on the correlation coefficient ρ of the population, which is unknown. This property, disadvantageous for a test of significance, is obviously characteristic for the nature of our problem. To demonstrate this dependence eleven different ρ-values covering the interval $-1 \leqq \rho \leqq +1$ were chosen, viz. the values $\rho = 0$, $\pm \cdot 2$, $\pm \cdot 4$, $\pm \cdot 6$, $\pm \cdot 8$, ± 1, and for these values were calculated the chances $P(\rho)$ (given by equation (1)) of obtaining two samples of 15 ($k = K = 6$) whose ratio of covariance estimates is greater than 74/26.*

The result is given below in Table I.

TABLE I

Giving the chance $P(\rho)$ of drawing two random samples of 15 from a normal population with correlation coefficient ρ such that the ratio of their covariance estimates is greater than 74/26

$\rho =$	0	$\pm \cdot 2$	$\pm \cdot 4$	$\pm \cdot 6$	$\pm \cdot 8$	± 1
$P(\rho) =$	·110	·128	·135	·097	·054	·030

From Table I it is obvious that there is no hope of approximating to our distribution (or to a transformation of it) by a normal curve (or any suitable distribution function) *independent of* ρ. Therefore, in testing significance we must admit all possible values of ρ, as we have started to do in Table I. Thus three different types of results may occur:

(*a*) For all values of ρ the P-values are smaller than ·05 (significant at 5 per cent.).

(*b*) For all values of ρ the P-values are greater than ·05 (insignificant at 5 per cent.).

(*c*) For some values of ρ the P-values are smaller than ·05, for other ρ-values the P-values are greater than ·05. The former ρ-values will cover an interval (or a set of intervals) on $-1 \leqq \rho \leqq +1$, which may be called I_1, the latter ρ-values will cover the remaining part of the range $-1 \leqq \rho \leqq +1$, which may be called I_2.

Table I shows an example of type (*c*), I_1 being (approximately) the intervals $\cdot 8 \leqq |\rho| \leqq 1$, I_2 being the interval $0 \leqq |\rho| \leqq \cdot 8$. To complete the test in cases like

* The value 74/26 has been chosen in connexion with a practical example to be discussed later.

this, the best estimate r of the correlation coefficient ρ has to be calculated from the samples, and the deviation of this observed value r from the nearest value in I_2, ρ' (say), taken as population value, has to be examined. In case this deviation is insignificant (at 5 per cent.), the whole test returns an insignificant result. For it has failed to disprove the hypothesis that both samples have been drawn from the same population with correlation coefficient ρ'. If, on the other hand, the deviation is significant, so is our original test. For, whatever the value of ρ may be, the original "null-hypothesis" has been disproved.

This latter part of the test is most easily worked out by Fisher's approximate method ((2), § 35) or by entering the exact distribution of r with ρ' as population value,* e.g. (1), p. xxxviii.

The first and main part of our test, however, consists in calculating the above chance $P(\rho)$ (see equations (1) and (2)), or rather in finding among all ρ-values, for which $P(\rho) \geqq \cdot 05$ (i.e. among all ρ-values in I_2, if any) that value ρ', which is nearest to our observed correlation coefficient.

To facilitate this, trivariate tables for the 5 per cent. points of the distribution of c (or a suitable transformation of it) would have to be worked out. The best arrangement of these would be in such a way, that two-way tables with the number of items in the larger sample as row-headings and (say) twenty different, positive ρ-values as column headings should proceed in pages with the number of items in the smaller sample.†

But even without the aid of such tables there is a method of working out the test for practical examples with the help of Pearson's tables of the incomplete B-function, and the calculations for such an example are shown below. Since Pearson's tables have been prepared so as to answer various other purposes, much calculation work is still left to be done, when applying them to our test. Though a further table (Table II) has been prepared, which facilitates the work considerably, the following method, unless both samples are very small, is still too laborious to become a common statistical practice.

PRACTICAL EXAMPLE

In a Cambridge nutrition experiment on pigs (3), among many other post-slaughter results the "mean back fat" (x) and the "percentage fat from back to belly" (y) were measured for 15 hogs and 15 gilts. One problem was to see how far mean back fat, which measurement is taken without great difficulty, provides a fair estimate of the percentage fat from back to belly and thus may be used for grading purposes.

We do not reproduce the 30 + 30 measurements here, but on examination it would be noted that the relationship was fairly marked for hogs but not so for

* In this case it might become necessary to perform the test for two ρ' values, viz. the nearest ρ-value in I_2 with $\rho < r$ (if any) and the nearest ρ-value in I_2 with $\rho > r$ (if any).

† The author regrets that, at the moment, he is unable to undertake this work, since he has only an adding machine at his disposal.

gilts. This sex difference is confirmed by the values of the respective sums of squares and products given below:

	(x^2)	(xy)	(y^2)	
Hogs	1·0975	10·488	236·06	......(3).
Gilts	0·6544	3·527	302·58	

Testing the significance of a correlation between y and x (or of a regression of y on x) for hogs and gilts separately a highly significant result is obtained for hogs whilst the gilts regression is quite insignificant. Nevertheless, we shall see that the difference between these relationships cannot be regarded as being significant. We shall test both the difference between correlation estimates (with the help of Fisher's z-transformation) and the difference between regression estimates (by the t-test), for in this example both questions are of interest. Finally, since variance estimates allow the assumption of equal population variances for hogs and gilts, we shall compare these tests with the test for the ratio of covariance estimates.

Let us start with the t-test. Doing this we obtain for b_H and b_G the respective values

$$b_H = 9{\cdot}56, \quad b_G = 5{\cdot}39.$$

Furthermore, the estimated S.D. of the difference $b_H - b_G$ has to be calculated in the usual way, the work being shown below:

$$\begin{aligned}
\text{Residual sum of squares (hogs)} &= 135{\cdot}83 \\
\text{,, \quad ,, \quad (gilts)} &= \underline{283{\cdot}57} \\
26 \times s^2 &= 419{\cdot}40 \\
s^2 &= 16{\cdot}131 \\
s^2/1{\cdot}0975 &= 14{\cdot}70 \\
s^2/\,{\cdot}6544 &= \underline{24{\cdot}65} \\
&\ 39{\cdot}35 \\
\text{S.D. of } (b_H - b_G) &= \sqrt{39{\cdot}35} \\
t &\sim {\cdot}665
\end{aligned}$$

The t-test returns an altogether insignificant difference; the chance of obtaining a difference equal to or larger than that observed being greater than 0·5.

Next we consider Fisher's test for the difference $z_1 - z_2$, i.e. the difference between the z-transformations of the estimated correlations r_1 and r_2. We obtain (approx.):

$$\left.\begin{aligned} r_1 = {\cdot}65, \quad z_1 = {\cdot}78 \\ r_2 = {\cdot}25, \quad z_2 = {\cdot}25 \end{aligned}\right\} \text{ diff.} = {\cdot}53;$$

S.D. of diff. $= {\cdot}408$;

normal deviate $= 1{\cdot}3$.

Again we obtain an insignificant result. The chance of obtaining a difference between correlation estimates larger than that observed is about 0·2.

The ratio of the covariance estimates, however, viz. that of the hogs divided by that of the gilts, is nearly 3, and since we know that our test approaches the z-test as ρ^2 approaches 1, this ratio will be significant for large ρ^2 The actual result of our test is summarized in Table I showing significance at 5 per cent. for $\cdot 8 \leqq |\rho| \leqq 1$ (approx.) and insignificance for $0 \leqq |\rho| \leqq \cdot 8$.*

Calculating now from equation (3) the best estimate of our correlation coefficient common to hogs and gilts, we obtain

$$r = \cdot 46.$$

This value lies right inside our I_2 interval and thus, having no deviation from the "nearest" ρ-value in I_2, returns our ratio c_0 as insignificant.

Though we obtain the same result as with the usual tests, it is obvious that in this example our test is more sensitive to the difference between the above covariance estimates. For our largest P-value is about 0·14 whilst the regression test yielded a P greater than 0·5 and Fisher's test of the z-transformation a P-value of about 0·2.

An explanation of the calculations, on which Table I is based, follows: We have to compute the value of $P^+(\rho)$ (given by equation (1)) for $k = K = 6$, $\rho = 0$, $\pm \cdot 2$, $\pm \cdot 4$, $\pm \cdot 6$, $\pm \cdot 8$, ± 1 and $c_0 = 74/26$.†

We first transform formula (1) into a form which is more suitable for its computation, whenever $k = K$.

Introducing the abbreviation

$$C(p, q; \rho) = 4^{-k-K-2} (kK)^{-1} \times 2^{p+q} [\mathrm{B}(k, k+2-q) \mathrm{B}(K, K+2-p)]^{-1} \times (1-\rho^2)^{k+K+2} \{(1-\rho)^{-p-q} + (1+\rho)^{-p-q}\},$$

we have
$$P^+(\rho) = \sum_{q=1}^{k+1} \sum_{p=1}^{K+1} I_x(p, q)\, C(p, q; \rho).$$

Now, since the I-function is only tabulated for $p \geqq q$, we write

$$P^+(\rho) = \sum_{q=1}^{k+1} \sum_{p=q}^{K+1} I_x(p, q)\, C(p, q; \rho) + \sum_{q=2}^{k+1} \sum_{p=1}^{q-1} I_x(p, q)\, C(p, q; \rho) \quad \ldots\ldots(4).$$

But since

$$I_x(p, q) = 1 - I_{1-x}(q, p),\ k = K \text{ and thus } C(p, q; \rho) = C(q, p; \rho),$$

we may write instead of equation (4)

$$P^+(\rho) = \sum_{q=1}^{k+1} \sum_{p=q}^{k+1} I_x(p, q)\, C(p, q; \rho) + \sum_{q=1}^{k} \sum_{p=q+1}^{k+1} \{1 - I_{1-x}(p, q)\}\, C(p, q; \rho),$$

* The value 74/26 shown in Table I is slightly smaller than the actual value of c_0 observed, viz. 2·974. This has been done to simplify the work and will be explained later.

† See the footnote to p. 68.

and finally we arrive at

$$P^{+}(\rho)=k^{-2}4^{-2k-2}(1-\rho^{2})^{2k+2}\left[\sum_{q=1}^{k+1}\frac{1}{\mathrm{B}(k,k+2-q)}\sum_{p=q}^{k+1}I_{x}(p,q)\frac{2^{p+q}}{\mathrm{B}(k,k+2-p)}\right.$$

$$\left.\times Q(\rho^{2},p+q)+\sum_{q=1}^{k}\frac{1}{\mathrm{B}(k,k+2-q)}\sum_{p=q+1}^{k+1}[1-I_{1-x}(p,q)]\frac{2^{p+q}}{\mathrm{B}(k,k+2-p)}Q(\rho^{2},p+q)\right], \qquad \ldots\ldots(5)$$

where $$Q(\rho^{2},p+q)=\{(1-\rho)^{-p-q}+(1+\rho)^{-p-q}\}.$$

These two sums are now most conveniently worked out together. The first step consists in preparing a table of the values

$$(1-\rho)^{-p-q}+(1+\rho)^{-p-q}.$$

This table, because independent of c_0, may be used by the reader for similar tests, provided the number of observations, when added for the two samples, is not greater than 30. The accuracy which is required for the entries of this table depends on the size of the sample, and (especially for high values of ρ) increases considerably as $p+q$ increases. Therefore a 5-figure table has been prepared rather than giving entries accurate up to a certain decimal.

TABLE II

Values of $(1-\rho)^{-p-q}+(1+\rho)^{-p-q}$

$(p+q)$	(*b*) $\rho=\cdot 2$	(*c*) $\rho=\cdot 4$	(*d*) $\rho=\cdot 6$	(*e*) $\rho=\cdot 8$
2	2·2569	3·2880	6·6406	$2{\cdot}5309\times 10$
3	2·5318	4·9941	$1{\cdot}5869\times 10$	$1{\cdot}2517\times 10^{2}$
4	2·9237	7·9764	$3{\cdot}9215^{+}\times 10$	$6{\cdot}2510\times 10^{2}$
5	3·4536	$1{\cdot}3046\times 10$	$9{\cdot}7752\times 10$	$3{\cdot}1251\times 10^{3}$
6	4·1496	$2{\cdot}1566\times 10$	$2{\cdot}4420\times 10^{2}$	$1{\cdot}5625\times 10^{4}$
7	$5{\cdot}0475^{-}$	$3{\cdot}5817\times 10$	$6{\cdot}1039\times 10^{2}$	5^{p+q}
8	6·1930	$5{\cdot}9605^{+}\times 10$	$1{\cdot}5259\times 10^{3}$	5^{p+q}
9	7·6444	$9{\cdot}9277\times 10$	$3{\cdot}8147\times 10^{3}$	5^{p+q}
10	9·4747	$1{\cdot}6542\times 10^{2}$	$9{\cdot}5368\times 10^{3}$	5^{p+q}
11	$1{\cdot}1776\times 10$	$2{\cdot}7566\times 10^{2}$	$2{\cdot}3842\times 10^{4}$	5^{p+q}
12	$1{\cdot}4664\times 10$	$4{\cdot}5941\times 10^{2}$	$5{\cdot}9605^{-}\times 10^{4}$	5^{p+q}
13	$1{\cdot}8283\times 10$	$7{\cdot}6567\times 10^{2}$	$1{\cdot}4901\times 10^{5}$	5^{p+q}
14	$2{\cdot}2815^{+}\times 10$	$1{\cdot}2761\times 10^{3}$	$3{\cdot}7253\times 10^{5}$	5^{p+q}

Next we write down the values of $[\mathrm{B}(6,8-q)]^{-1}$ for $q=1, \ldots, 7$, which are easily obtained from their definition, viz.

$$[\mathrm{B}(6,8-q)]^{-1}=(13-q)!/(7-q)!\,5!$$

Finally, we prepare with the help of Pearson's tables of the incomplete B-function a table of the values

$$2^{p+q}[\mathrm{B}(6,8-p)]^{-1}\{I_{\cdot 26}(p,q)+1-I_{\cdot 74}(p,q)\} \text{ for } p>q.$$

TABLE III

Values of $1/\mathrm{B}(6, 8-q)$

q	1	2	3	4	5	6	7
$1/\mathrm{B}(6, 8-q)$	5544	2772	1260	504	168	42	6

In choosing ·26 for X we have substituted for c_0 a value which is slightly smaller than the ratio actually observed, in order to coincide with an entry in Pearson's table. This we have done to save the reader the work of about 56 interpolations. The diagonal entries of the table are the values $2^{2p}[\mathrm{B}(6, 8-p)]^{-1} \times I_{\cdot 26}(p, p)$, since these are required for the further calculations.

TABLE IV

Values of $\dfrac{2^{p+q}}{\mathrm{B}(6, 8-p)}\{I_{\cdot 26}(p, q) + 1 - I_{\cdot 74}(p, q)\}$ *for* $p > q$ *and* $\dfrac{2^{p+q}}{\mathrm{B}(6, 8-p)} I_{\cdot 26}(p, p)$ for $p = q$

p \ q	1	2	3	4	5	6	7	(*)
1	5,765·8							
2	11,531·5	7,435·5						
3	12,345·0	13,519·1	9,220·6					
4	11,365·5	13,107·9	14,752·9	10,342·0				
5	8,378·9	10,649·1	12,040·7	13,789·4	9,829·0			
6	4,494·9	6,250·3	7,564·2	8,478·8	9,829·0	7,085·8		
7	1,349·5	2,020·9	2,636·2	3,101·5	3,461·4	4,049·0	2,943·2	
(*a*) $\rho = 0$	$11{,}046 \times 10$	$10{,}597 \times 10$	$9{,}243 \times 10$	$7{,}142 \times 10$	$4{,}624 \times 10$	$2{,}227 \times 10$	5,886	$1{,}067 \times$
(*b*) $\rho = \cdot 2$	$1{,}834 \times 10^2$	$2{,}307 \times 10^2$	$2{,}701 \times 10^2$	$2{,}863 \times 10^2$	260×10^3	178×10^3	67×10^3	$219 \times$
(*c*) $\rho = \cdot 4$	75×10^4	$1{,}473 \times 10^3$	$2{,}632 \times 10^3$	$4{,}243 \times 10^3$	593×10^4	636×10^4	38×10^5	$150 \times$
(*d*) $\rho = \cdot 6$	87×10^5	$2{,}856 \times 10^4$	836×10^5	$2{,}232 \times 10^5$	534×10^6	$1{,}026 \times 10^6$	110×10^7	$48 \times$
(*e*) $\rho = \cdot 8$	105×10^7	75×10^8	465×10^8	265×10^9	$1{,}421 \times 10^9$	$6{,}673 \times 10^9$	$1{,}796 \times 10^{10}$	$85 \times$

The rest of the work is obvious from formula (5) and is shown in the above table.†

In rows (*a*), (*b*), (*c*), (*d*) and (*e*) are given weighted totals of the respective columns in the upper part of Table IV. To calculate the weighted column totals, which are shown in row (*b*) (say) we multiply the entry in the pth row and in the qth column of Table IV by that entry of Table II which is shown in the $(p+q)$th row and in the column (*b*). Then these products are summed to yield the weighted column totals in row (*b*) of Table IV. Similarly, for the rows (*c*), (*d*) and (*e*). The entries of row (*a*) are simply obtained by doubling the column totals of Table IV.

Finally the qth entry of each of the rows (*a*), (*b*), (*c*), (*d*) and (*e*) in Table IV is multiplied by the qth entry of Table III and these products summed for each row,

† The great loss of accuracy is due to the small scope of an ordinary adding machine, with which the work had to be done.

the respective sums of products being shown in column (*). If these sums of products are divided by the respective values of $36 \times 4^{14} \times (1-\rho^2)^{-14}$, the chances $P^+(\rho)$ of Table I are obtained.

My heartiest thanks are due to Dr J. Wishart for his help and advice throughout this work and to Mr F. J. Dudley for his help in preparing Table II.

APPENDIX

Derivation of the distribution function for the ratio of covariance estimates

(i) In this part we shall give a mathematical proof of the formulae (1) and (2), which have been used for a practical test in the preceding part of the paper. Incidentally we shall derive the theory for a more general problem and point out further properties of our distribution, which are analogous to the well-known behaviour of the z-distribution.

The problem may be stated in its generalized form straightaway:

Let $\qquad x_i, y_i \qquad (i=1, 2, \ldots, n'=2k+3)$

and $\qquad X_j, Y_j \qquad (j=1, 2, \ldots, N'=2K+3)$

be two (independent) random samples drawn from the populations

$$f(x, y) = [(\alpha\beta - \nu^2)^{\frac{1}{2}}/\pi] \times \exp\{-(\alpha x^2 + 2\nu xy + \beta y^2)\} \qquad \ldots\ldots(6)$$

and

$$F(X, Y) = [(\mathrm{AB} - \mathrm{N}^2)^{\frac{1}{2}}/\pi] \times \exp\{-(\mathrm{A}X^2 + 2\mathrm{N}XY + \mathrm{B}Y^2)\} \qquad \ldots\ldots(7)$$

respectively.†

It is then required to obtain the distribution function for the ratio of covariance estimates

$$c = [(N'-1)/(n'-1)] \times \sum_{i=1}^{n'} (x_i - \bar{x})(y_i - \bar{y}) \Big/ \sum_{j=1}^{N'} (X_j - \bar{X})(Y_j - \bar{Y}),$$

where by $\bar{x}$, $\bar{y}$, $\bar{X}$, $\bar{Y}$ we denote the respective arithmetic means of the samples, viz.

$$\frac{1}{n'}\sum_{i=1}^{n'} x_i, \quad \frac{1}{n'}\sum_{i=1}^{n'} y_i, \quad \frac{1}{N'}\sum_{j=1}^{N'} X_j, \quad \frac{1}{N'}\sum_{j=1}^{N'} Y_j.$$

In the course of the proof, the ratio of the "sums of products",

$$w = \sum_{i=1}^{n'} (x_i - \bar{x})(y_i - \bar{y}) \Big/ \sum_{j=1}^{N'} (X_j - \bar{X})(Y_j - \bar{Y}), \qquad \ldots\ldots(8)$$

will turn out to be a more convenient statistic than c itself.

† Instead of the parameters α, β, ν the quantities σ_1, σ_2, ρ (i.e. the standard deviations of x and of y and the correlation between x and y) are more commonly used to represent a normal population. In terms of σ_1, σ_2, ρ the parameters α, β, ν are defined by

$$\alpha^{-1} = 2\sigma_1^2(1-\rho^2), \quad \beta^{-1} = 2\sigma_2^2(1-\rho^2), \quad \nu = -\rho/\{2\sigma_1\sigma_2(1-\rho^2)\}.$$

(ii) Our first step is to show that the distribution of the sum of products

$$u=\sum_{i=1}^{n'}(x_i-\bar{x})(y_i-\bar{y}) \qquad \ldots\ldots(9)$$

may be expressed as a finite sum of elementary functions, provided n', the number of items in the sample, is odd. If we introduce new variates ξ, η by the linear transformation

$$\left.\begin{aligned} x&=\sqrt{\beta/\alpha}\,\xi+\eta \\ y&=\xi-\sqrt{\alpha/\beta}\,\eta \end{aligned}\right\} \qquad \ldots\ldots(10)$$

then by substituting (10) in (6) we obtain for the distribution of ξ, η

$$[2(\alpha\beta-\nu^2)^{\frac{1}{2}}/\pi]\times\exp\{-(2\beta+2\nu\sqrt{\beta/\alpha})\,\xi^2-(2\alpha-2\nu\sqrt{\alpha/\beta})\,\eta^2\}. \quad \ldots\ldots(11)$$

Now to any sample of n' pairs x_i, y_i there will correspond a sample of n pairs ξ_i, η_i obtained by the converse of (10). Furthermore, obviously

$$u=\sqrt{\beta/\alpha}\,\gamma_{11}-\sqrt{\alpha/\beta}\,\gamma_{22},$$

if $$\gamma_{11}=\sum_{i=1}^{n'}(\xi_i-\bar{\xi})^2,\quad \gamma_{22}=\sum_{i=1}^{n'}(\eta_i-\bar{\eta})^2,\quad \bar{\xi}=\left(\frac{1}{n'}\right)\sum_{i=1}^{n'}\xi_i,\quad \bar{\eta}=\left(\frac{1}{n'}\right)\sum_{i=1}^{n'}\eta_i. \qquad \ldots\ldots(12)$$

From equation (11) it is obvious that the variates ξ, η are independently distributed. Hence the joint distribution of γ_{11}, γ_{22} (see e.g. (4)) is given by

$$\begin{aligned}\psi(\gamma_{11},\gamma_{22})=4^{k+1}(\alpha\beta-\nu^2)^{k+1}(k\,!)^{-2}\gamma_{11}^k\gamma_{22}^k \\ \times\exp\{-(2\beta+2\nu\sqrt{\beta/\alpha})\,\gamma_{11}-(2\alpha-2\nu\sqrt{\alpha/\beta})\,\gamma_{22}\},\end{aligned}$$

where $k=(n'-3)/2$ is an integer $\geqslant 1$. Thus by equation (12) we arrive at the joint distribution of u and γ_{22}, viz.

$$\begin{aligned}\chi(u,\gamma_{22})=4^{k+1}(\alpha\beta-\nu^2)^{k+1}(k\,!)^{-2}(\sqrt{\alpha/\beta})^{k+1}\exp\{-(2\sqrt{\alpha\beta}+2\nu)\,u\} \\ \times(u+\sqrt{\alpha/\beta}\,\gamma_{22})^k\gamma_{22}^k\exp\{-4\alpha\gamma_{22}\}. \quad \ldots\ldots(13)\end{aligned}$$

To obtain the distribution function $g(u)$ (say) of the sum of products we have to integrate $\chi(u,\gamma_{22})$ (see equation (13)) over the range of γ_{22}. This range, however, depends on u. For since the range of the variates γ_{11}, γ_{22} is given by

$$0\leqq\gamma_{11}<\infty,\quad 0\leqq\gamma_{22}<\infty,$$

the range for u, γ_{22} must be

$$0\leqq\gamma_{22}<\infty,\quad -u\sqrt{\beta/\alpha}\leqq\gamma_{22},\quad -\infty<u<+\infty.$$

We first consider the case $\qquad u\geqq 0.$

In this case we have to integrate $\chi(u,\gamma_{22})$ for γ_{22} ranging from 0 to $+\infty$. To do this

we apply partial integration $(2k+1)$ times, integrating $\exp\{-4\alpha\gamma_{22}\}$, differentiating the powers of γ_{22} and collecting the terms at $\gamma_{22}=0$. We thus obtain

$$g^{+}(u)=4^{k+1}(\alpha\beta-\nu^2)^{k+1}(k\,!)^{-2}(\sqrt{\alpha/\beta})^{k+1}\exp\{-(2\sqrt{\alpha\beta}+2\nu)u\}$$
$$\times\sum_{\tau=0}^{k}\frac{(k+\tau)!\,k!}{(k-\tau)!\,\tau!}(\sqrt{\alpha/\beta})^{\tau}u^{k-\tau}(4\alpha)^{-k-\tau-1},$$

which may be written as

$$g^{+}(u)=(\alpha\beta-\nu^2)^{k+1}\exp\{-(2\sqrt{\alpha\beta}+2\nu)u\}(\sqrt{\alpha\beta})^{-k-1}$$
$$\times\sum_{\tau=0}^{k}\frac{(k+\tau)!\,(4\sqrt{\alpha\beta})^{-\tau}}{k!\,\tau!\,(k-\tau)!}u^{k-\tau}. \quad\ldots\ldots(14)$$

We next consider the case $\qquad u\leqq 0.$

Now we have to integrate for γ_{22} ranging from $-u\sqrt{\beta/\alpha}$ to $+\infty$. Integrating by parts as above and collecting the terms at $\gamma_{22}=-u\sqrt{\beta/\alpha}$ we have

$$g^{-}(u)=(\alpha\beta-\nu^2)^{k+1}\exp\{(2\sqrt{\alpha\beta}-2\nu)u\}(\sqrt{\alpha\beta})^{-k-1}$$
$$\times\sum_{\tau=0}^{k}\frac{(k+\tau)!\,(4\sqrt{\alpha\beta})^{-\tau}}{k!\,\tau!\,(k-\tau)!}(-u)^{k-\tau}. \quad\ldots\ldots(15)$$

The required distribution function for u is thus given by

$$g(u)=\begin{cases} g^{+}(u) & \text{for}\quad u\geqq 0\quad(\text{see equation (14)})\\ g^{-}(u) & \text{for}\quad u\leqq 0\quad(\text{see equation (15)})\end{cases}. \qquad\ldots\ldots(16)$$

Comparing these with the "Bessel-function distribution" of $2\sqrt{\alpha\beta}\,u$ (5), (4), we incidentally have proved the well-known fact that the Bessel-function $K_{\nu}(x)$, for fractional ν, can be expressed by elementary functions in a finite form.

(iii) It is now easy to see how the distribution of the ratio of the two sums of products

$$w=\frac{u}{U}=c\,[(k+1)/(K+1)]$$

can be derived from equations (14) and (15) by elementary integrations. For if we consider the sum of products

$$U=\sum_{j=1}^{2K+3}(X_j-\bar{X})(Y_j-\bar{Y}),$$

the sample of which has been drawn from the population (7), then the distribution of U (say $G(U)$) is obtained by replacing in equations (14), (15) and (16) the letters u, α, β, ν, g, $k=(n'-3)/2$ by U, A, B, N, G, $K=(N'-3)/2$, respectively. Hence, independence of the samples (x_i, y_i) and (X_j, Y_j) being assumed, the joint distribution of u and U is equal to $g(u)\times G(U)$ and thus the distribution function of w (say $\Phi(w)$) is given by

$$\Phi(w)=\int_{-\infty}^{+\infty}g(wU)\,G(U)\,|\,U\,|\,dU, \qquad\ldots\ldots(17)$$

since the modulus of the Jacobian of the transformation

$$u = wU, \quad U = U$$

is simply equal to $|U|$. To work out equation (17) we first consider the case

$$w \geqq 0.$$

We now may write

$$\Phi^+(w) = -\int_{-\infty}^{0} g^-(wU)\, G^-(U)\, U\, dU + \int_0^\infty g^+(wU)\, G^+(U)\, U\, dU$$
$$= T_1(w) \qquad\qquad + T_2(w) \text{ (say)},$$

and start to consider $T_2(w)$. Now by equation (14) plainly

$$T_2(w) = (\alpha\beta - \nu^2)^{k+1} (\sqrt{\alpha\beta})^{-k-1} (\mathrm{AB} - \mathrm{N}^2)^{K+1} (\sqrt{\mathrm{AB}})^{-K-1} (k!\, K!)^{-1}$$
$$\times \sum_{\tau=0}^{k} \frac{(k+\tau)!}{(k-\tau)!\, \tau!} \sum_{\mu=0}^{K} \frac{(K+\mu)!}{(K-\mu)!\, \mu!} (4\sqrt{\alpha\beta})^{-\tau} (4\sqrt{\mathrm{AB}})^{-\mu}$$
$$\times w^{k-\tau} \int_0^\infty U^{k+K-\tau-\mu+1} \exp\{-2[(\sqrt{\alpha\beta}+\nu)w + (\sqrt{\mathrm{AB}}+\mathrm{N})]\, U\}\, dU,$$

where the last integral, by partial integration, is seen to be equal to

$$(k+K-\tau-\mu+1)!\, [2\{(\sqrt{\alpha\beta}+\nu)w + (\sqrt{\mathrm{AB}}+\mathrm{N})\}]^{-k-K+\tau+\mu-2}.$$

If then $T_1(w)$ is worked out in the same way with the help of equation (15) we finally have

$$\Phi^+(w) = [(\alpha\beta - \nu^2)/\sqrt{\alpha\beta}]^{k+1} [(\mathrm{AB} - \mathrm{N}^2)/\sqrt{\mathrm{AB}}]^{K+1} (k!\, K!)^{-1}$$
$$\times \sum_{\tau=0}^{k} \frac{(k+\tau)!}{(k-\tau)!\, \tau!} \sum_{\mu=0}^{K} \frac{(K+\mu)!}{(K-\mu)!\, \mu!} (2\sqrt{\alpha\beta})^{-\tau} (2\sqrt{\mathrm{AB}})^{-\mu}\, 2^{-k-K-2}$$
$$\times (k+K-\tau-\mu+1)!\, w^{k-\tau} \{[(\sqrt{\alpha\beta}+\nu)w + (\sqrt{\mathrm{AB}}+\mathrm{N})]^{-k-K+\tau+\mu-2}$$
$$+ [(\sqrt{\alpha\beta}-\nu)w + (\sqrt{\mathrm{AB}}-\mathrm{N})]^{-k-K+\tau+\mu-2}\}, \quad \ldots\ldots(18)$$

or, introducing $i = k-\tau$ and $j = K-\mu$ as summation indices,

$$\Phi^+(w) = [(\alpha\beta - \nu^2)/\sqrt{\alpha\beta}]^{k+1} [(\mathrm{AB} - \mathrm{N}^2)/\sqrt{\mathrm{AB}}]^{K+1} (k!\, K!)^{-1}$$
$$\times \sum_{i=0}^{k} \frac{(2k-i)!}{i!\,(k-i)!} \sum_{j=0}^{K} \frac{(2K-j)!}{j!\,(K-j)!} \frac{(2\sqrt{\alpha\beta})^{-k+i} (2\sqrt{\mathrm{AB}})^{-K+j}}{2^{k+K+2}} (i+j+1)!$$
$$\times w^i \{[(\sqrt{\alpha\beta}+\nu)w + (\sqrt{\mathrm{AB}}+\mathrm{N})]^{-i-j-2} + [(\sqrt{\alpha\beta}-\nu)w + (\sqrt{\mathrm{AB}}-\mathrm{N})]^{-i-j-2}\}.$$
$$\ldots\ldots(19)$$

Turning now to the case $\qquad w \leqq 0$

we obtain by the same argument as above for

$$\Phi^-(w) = \int_{-\infty}^{0} g^+(wU)\, G^-(U)\, |U|\, dU + \int_0^\infty g^-(wU)\, G^+(U)\, U\, dU$$

the sum

$$\Phi^-(w)=[(\alpha\beta-\nu^2)/\sqrt{\alpha\beta}]^{k+1}[(\mathrm{AB}-\mathrm{N}^2)/\sqrt{\mathrm{AB}}]^{K+1}(k!\,K!)^{-1}$$
$$\times\sum_{i=0}^{k}\frac{(2k-i)!}{i!\,(k-i)!}\sum_{j=0}^{K}\frac{(2K-j)!}{j!\,(K-j)!}\frac{(2\sqrt{\alpha\beta})^{-k+i}(2\sqrt{\mathrm{AB}})^{-K+j}}{2^{k+K+2}}(i+j+1)!$$
$$\times(-w)^i\{[(\sqrt{\alpha\beta}+\nu)(-w)+(\sqrt{\mathrm{AB}}-\mathrm{N})]^{-i-j-2}$$
$$+[(\sqrt{\alpha\beta}-\nu)(-w)+(\sqrt{\mathrm{AB}}+\mathrm{N})]^{-i-j-2}\}. \quad ......(20)$$

(iv) To use equations (19) and (20) for a test of significance we confine ourselves to the most important special case, viz. we assume that the populations (6) and (7) are identical, whence

$$\alpha=\mathrm{A},\quad \beta=\mathrm{B},\quad \nu=\mathrm{N}.^* \quad ......(21)$$

We first consider the distribution $\Phi(w)$ for

$$w\geqq 0.$$

Using equation (21) and introducing the correlation coefficient

$$\rho=-\nu/\sqrt{\alpha\beta},$$

equation (19) may be written as

$$\Phi^+(w)=\sum_{i=0}^{k}\sum_{j=0}^{K}\frac{(2k-i)!}{(k-i)!\,k!}\frac{(2K-j)!}{(K-j)!\,K!}\frac{(i+j+1)!}{i!\,j!}\frac{w^i}{(1+w)^{i+j+2}}$$
$$\times 2^{-2k-2K+i+j-2}(1-\rho^2)^{k+K-i-j}\{(1-\rho)^{i+j+2}+(1+\rho)^{i+j+2}\}. \quad ......(22)$$

Thus $\Phi^+(w)$ is seen to depend on $k=(n'-3)/2$, $K=(N'-3)/2$ and ρ^2. Furthermore, as $\rho^2\to 1$,

$$\Phi^+(w)\to\frac{(k+K+1)!}{k!\,K!}\frac{w^k}{(1+w)^{k+K+2}},$$

and this is the distribution function for $w=[(k+1)/(K+1)]\,e^{2z}$.

To prove formula (1) we have to determine the probability integral

$$P^+=\int_W^\infty\Phi^+(w)\,dw\quad\text{for any } W\geqq 0.$$

But since
$$\int_W^\infty\frac{w^i}{(1+w)^{i+j+2}}dw=-\int_{1/(1+W)}^{0}\omega^{i+j+2}(1-\omega)^i\omega^{-i-2}\,d\omega$$
$$=\int_0^{1/(1+W)}\omega^j(1-\omega)^i\,d\omega,$$

we see from equation (22) that we can express our probability integral with the help of the incomplete B-functions. Using the notation of p. 67 we obtain

$$P^+=\int_W^\infty\Phi^+(w)\,dw=4^{-k-K-2}k^{-1}K^{-1}$$
$$\times\sum_{q=1}^{k+1}\sum_{p=1}^{K+1}I_{[1/1+W]}(p,q)\frac{2^{q+p}}{\mathrm{B}(k,k+2-q)\,\mathrm{B}(K,K+2-p)}$$
$$\times(1-\rho^2)^{k+K+2-p-q}\{(1-\rho)^{p+q}+(1+\rho)^{p+q}\}, \quad ......(23)$$

which is equivalent to formula (1).

* In order to derive formulae (1) and (2) it would be sufficient to assume that for the populations (6) and (7) the equations $\nu=\mathrm{N}$ and $\alpha\beta=\mathrm{AB}$ are fulfilled.

A further remark may be added:

If the two independent samples had been drawn from populations having the same correlation coefficient but unequal variances (i.e. if in equations (6) and (7) $\nu/\sqrt{\alpha\beta} = \mathrm{N}/\sqrt{\mathrm{AB}} = -\rho$), then it is easy to see that again test (23) is valid with w and W replaced by $w' = w\,(\sqrt{\alpha\beta}/\sqrt{\mathrm{AB}})$ and $W' = W(\sqrt{\alpha\beta}/\sqrt{\mathrm{AB}})$ respectively.* This is analogous to the well-known property of the e^{2z} distribution.

We now turn to the case of negative ratios w:

$$w \leqq 0.$$

From equation (20) we obtain under assumption (21) for the distribution of negative values of w

$$\Phi^-(w) = \sum_{i=0}^{k} \sum_{j=0}^{K} \frac{(2k-i)!}{(k-i)!\,k!} \frac{(2K-j)!}{(K-j)!\,K!} \frac{(i+j+1)!}{i!\,j!} 2^{-2k-2K+i+j-2}$$
$$\times \left[\frac{(-w)^i}{[(1+\rho)(-w)+(1-\rho)]^{i+j+2}} + \frac{(-w)^i}{[(1-\rho)(-w)+(1+\rho)]^{i+j+2}} \right]$$
$$\times (1-\rho^2)^{k+K+2}. \qquad \ldots\ldots(24)$$

It is easy to see that, as $\rho^2 \to 1$, $\Phi^-(w) \to 0$ uniformly in any finite interval of non-positive w-values.

Finally, we obtain by elementary transformations for any $W \geqq 0$

$$P^- = \int_{-\infty}^{-W} \Phi^-(w)\,dw = 4^{-k-K-2}(kK)^{-1} \times \sum_{q=1}^{k+1} \sum_{p=1}^{K+1} \frac{2^{p+q}}{\mathrm{B}(k, k+2-q)\,\mathrm{B}(K, K+2-p)}$$
$$\times \left[(1-\rho)^{-p}(1+\rho)^{-q} I_{\left[1\Big/\left(1+W\frac{1+\rho}{1-\rho}\right)\right]}(p,q) \right.$$
$$\left. + (1+\rho)^{-p}(1-\rho)^{-q} I_{\left[1\Big/\left(1+W\frac{1-\rho}{1+\rho}\right)\right]}(p,q) \right] \times (1-\rho^2)^{k+K+2}, \quad \ldots\ldots(25)$$

which is equivalent to formula (2).

If the two independent samples had been drawn from populations with different variances but equal correlation coefficients, then what has been stated about P^+ also applies to P^-.

* Thus the calculation of $P^+(\rho)$ only differs from that described previously in that a different row of Pearson's table has to be entered. Should it, however, become necessary to calculate the best estimate of ρ, it has to be remembered that now the populations (6) and (7) may have different variances.

REFERENCES

(1) PEARSON, K. *Tables of the Incomplete Beta-function* (1934). (Biometric Laboratory, Univ. Coll. London.)

(2) FISHER, R. A. *Statistical Methods for Research Workers*, 5th ed. (1934).

(3) WOODMAN, H. E., EVANS, R. E., CALLOW, E. H. and WISHART, J. "The nutrition of the bacon pig: (1) The influence of high levels of protein-intake on growth, conformation and quality of the bacon pig." *J. Agric. Sci.* XXVI (1936), pp. 546–619.

(4) WISHART, J. and BARTLETT, M. S. *Proc. Camb. Phil. Soc.* XXVIII (1932), pp. 455–9.

(5) PEARSON, K., JEFFERY, G. B. and ELDERTON, E. M. *Biometrika*, XXI (1929), pp. 164–93.

(6) PEARSON, E. S. and WILKS, S. S. *Biometrika*, XXV (1933), pp. 353–78.

A CONTRIBUTION TO THE BIOMETRIC STUDY OF THE HUMAN MANDIBLE

By FRANK H. CLEAVER, M.A.

Crewdson Benington Student in Craniometry

1. *Introduction.* It is clear to-day that the statistical study of anthropometric data with the object of investigating racial origins and relationships requires numbers of subjects, or specimens, far in excess of those considered sufficient by earlier anthropologists. The newer methods also demand greater precision in measurement, and better control and standardization of the techniques used in collecting the data. There is far more metrical material available for the cranium than for any other part of the skeleton, and the results for it are far in advance of those for any other kind of anthropometric material. The available measurements of living series, though more extensive, are unfortunately of lesser value owing to the fact that there has been no effective control or standardization of the techniques used in determining them. Owing largely to recent work presented in papers in *Biometrika*, the mandible is now the part of the skeleton, after the cranium, which has been best described metrically. The present paper provides statistical data for four new male and three corresponding female series, viz. two English, a Punjabi (male only) and an Australian. In all there are now 17 male and 9 female series measured by following the same biometric technique, though it is clear that some of these are too small to be of permanent value by themselves. The statistical treatment applied here is the same as that of the earlier papers, and particular attention is paid to a discussion of the coefficients of racial likeness. It was not to be expected that the conditions which have to be fulfilled in interpreting these criteria of resemblance would be *precisely* the same for mandibular as for cranial material. There are clear differences between the two kinds of evidence in this respect, and a general conclusion reached is that measurements of more and longer series of mandibles are still needed in order to discover how far the bone is capable of revealing racial relationships. It appears to be less effective for this purpose than the cranium.

2. *Description of the material.* Original data relating to four series of mandibles are presented in this paper. The material consists of two London series (from Spitalfields and Farringdon Street), a Punjabi and a native Australian. The first two of these are preserved at University College, London, and permission to work on them was kindly granted by the late Professor Karl Pearson; the third and fourth, preserved in the Museum of the Royal College of Surgeons of England, were loaned by the College authorities, and I should like to take this opportunity

of thanking them for their kindness, and the staff of the College for the consideration and assistance I at all times received from them.

(*a*) The Spitalfields skeletons were dug up in 1926,* and the nature of the interment in roughly circular pits, without any orderly arrangement of the bodies, pointed to a mass burial, the result of plague, massacre, or some such catastrophe. Examination of the crania showed that they are racially homogeneous, while comparisons between the crania of this and of other series (using the method of the coefficient of racial likeness) have indicated that the Spitalfields series lies closer to Pompeians, Etruscans and the population interred in the Church of St Leonard, Hythe, than it does to any other European series available. The type is far removed from those of the Neolithic, Bronze Age and Anglo-Saxon populations of England, as well as from seventeenth-century crania excavated at Whitechapel and Farringdon Street. The Spitalfields interment is therefore one of an intrusive population, and in the absence of datable artifacts, and of any other direct archaeological evidence, it has been concluded that it took place either in mediaeval or Roman times, this assumption being based mainly on an examination of the history of the Spitalfields site. The measured series is made up by 63 male and 32 female adult mandibles, only 12 of these being associated with crania. There are 195 other mandibles from Spitalfields which were not measured, either because they are immature, or else because they are too fragmentary for the purpose. Nearly 1000 crania from the site were preserved—the majority of these being incomplete—and it is estimated that about 3000 people were buried in the excavated area.

(*b*) The Farringdon Street skeletons, of which the mandibles discussed in the present paper form part, were dug up in 1924. A detailed account of the evidence for dating the bones was prepared by Professor Karl Pearson.† As a result of his examination it is safe to say that the interment of the Farringdon Street skeletons took place during the period 1610–1722 in the graveyard of the Parish Church of St Bride, but that the majority of the interments were made between 1610 and 1666 and were mainly the results of deaths from the Great Plague, 1665. Miss Beatrix G. E. Hooke measured over 350 of the Farringdon Street crania and 67 of the mandibles. The measurements of the unsexed mandibles were published in her paper, "A Third Study of the English Skull with special reference to the Farringdon Street Crania".‡ She states that several hundred mandibles were dug up, none being attached to skulls, but that the incomplete condition of these bones, due to breakages, prevented the taking of a fairly complete set of measurements except on 67. The present writer re-examined the collection of Farringdon Street mandibles and, bearing in mind certain requirements necessary for purposes of sexing (a more detailed account of which will be found in another section

* G. M. Morant and M. F. Hoadley, "A Study of the Recently Excavated Spitalfields Crania", *Biometrika*, XXIII (1931), pp. 191–248.

† *Biometrika*, XVIII (1926), pp. 1–15.

‡ *Ibid.* pp. 1–55.

of this paper), he was able to pick out and measure 90, i.e. 23 more than Miss Hooke measured. This difference can possibly be ascribed to the fact that Miss Hooke chose for measurement only those mandibles on which all, or nearly all, the 35 measurements used at that time could be taken, while the present writer used only 16 of those measurements selected as being the most reliable (see section 3 below).

(*c*) The Australian mandibles dealt with in this paper are those in the Museum collections of the Royal College of Surgeons. These specimens were obtained from several sources at different times, and there are no series of any length among them from single burial-grounds. They are divided in the Museum catalogue according to a scheme of classification based on the modern State territories, but for purposes of statistical treatment they have been pooled, with the exception of five from the Northern Territory. The justification for this procedure is that a statistical examination of about 300 Australian crania collected from all over the continent suggests that only two racial divisions can be recognized—that from the Northern Territory, where immigration is most likely to have affected the type, and that which is spread over the enormous area of the rest of Australia.* There is, as would be expected, a close connection between these two groups. An examination of the cranial facial skeleton of the two racial groups distinguished above revealed no significant differences between them, and hence it would have been of considerable interest to compare the Northern Territory mandibles with those from the remaining area. Unfortunately, no adequate numbers were available for the former group which was demarcated from the remainder in accordance with the cranial evidence mentioned above. Of this remainder the males were first divided into two sets. The first comprises those from Western Australia (3), South Australia (17) and Victoria (16), and the second those from Queensland (14) and New South Wales (9). This was a purely arbitrary division based on geographical position, and a comparison of the two groups revealed no significant differences at all between them.† Pooling is hence justified as far as can be seen from this evidence. Nine male mandibles from unknown localities were included in the pooled series. There are 36 female adult Australian mandibles: 2 from Western Australia, 12 from South Australia, 4 from Victoria, 9 from Queensland, 6 from New South Wales and 3 from unknown localities.

(*d*) The series described as Punjabi in this paper comprises those mandibles catalogued as such at the Royal College of Surgeons, and it is made up almost entirely of male mandibles from the collection which Sir Havelock Charles

* G. M. Morant, "A Study of the Australian and Tasmanian Skulls, based on previously published Measurements", *Biometrika*, XIX (1927), pp. 417–40.

† The coefficient of racial likeness based on cranial measurements for these two groups is $1{\cdot}58 \pm {\cdot}21$ for the male and ${\cdot}69 \pm {\cdot}21$ for the female series (Morant, *loc. cit.* p. 424) No great reliance can be placed on any generalizations concerning the racial composition of the whole of the Australian continent in view of the scanty nature of the material available

presented to the Museum of the College. They belonged for the most part to inmates of the British Hospital at Lahore, where Sir Havelock Charles was a surgeon, and as such they cannot really be considered as a random sample of the population of the Punjab. The few other mandibles included in the series are said to have come from various parts of the Punjab, and the specimens of the whole collection are variously catalogued as Sikh, Jat, Pathan, etc., etc. The main scheme of classification, however, distinguishes two groups—Hindu and Mohammédan—the basis of distinction being thus religious and not ethnological. In the Punjab, besides Sikhs and Pathan immigrants from across the frontier—both Mohammedan conquering stocks—there are the Mohammedan converts. The religion of Islam seems to have taken a firm hold on the native population, and, judging from census returns, large numbers of Jats, Rajputs and Gujars were of the Mohammedan faith. Just as it is confidently asserted that in Bengal the Mohammedans are of the same racial type as the lowest castes of Hindus, so in the Punjab the former are not clearly distinguished from the Hindus, though the religious divisions appear to be of significance from a racial point of view as will be shown below. The British Hospital at Lahore would no doubt have admitted the Sikhs, the Pathans, the descendants of the old Rajput rulers, the Jat peasantry and perhaps even some of the nomad Baloches, who are supposed to be of a distinctive physical type. The present sample cannot be considered a racially homogeneous one, and it must be considered from the racial standpoint only as representing an Indian type, in contrast to the European and Australian types also dealt with in this paper. Such lack of homogeneity was evidenced when the sample itself was divided into two groups—Mohammedan and Hindu. A comparison for all characters between the male groups—made up by 27 and 22 mandibles respectively—shows that differences exceeding 3·5 times their probable errors are found only for ml (Δ/(p.e. Δ) = 6·4) and $C' \angle$ (4·6) out of the 21 characters compared. The coefficient of racial likeness between these two series was calculated for 10 characters,* giving a crude value of 1·61 ± ·30 and a reduced of 7·47 ± 1·40. There is no doubt that the total Punjabi series is racially heterogeneous, but division by religion may not be the best possible, and for practical purposes it seemed advisable to use the total group, which is far from an ideal procedure. The fact that in this paper the pooled sample has been treated like a homogeneous series may be partly responsible for the unsatisfactory results (given below) which are found when racial comparisons are made between this series from the Punjab and the Farringdon Street and other series. In all 49 male adult, 9 female adult and 2 immature Punjabi mandibles were measured.

A list of all the previous series of mandibles on which measurements have been taken in accordance with the biometric technique is given by G. M. Morant.†

* A list of the characters used is given in a footnote to Table X and the Qau Egyptian standard deviations were used in calculating the coefficient.

† *Biometrika*, XXVIII (1936), pp. 92–4.

He refers to 2 European, 4 Egyptian, and 6 Asiatic series. E. S. Martin has provided similar data for another ancient Egyptian series.* Comparisons between the four series dealt with for the first time in the present paper and the earlier material are made below.†

3. *Definitions of measurements and estimates of their accuracy.* The biometric technique for measuring the human mandible was given by G. M. Morant in *Biometrika*, XIV, pp. 253–60, in 1923. Only those measurements included in a revised list,‡ and chosen chiefly because they could be taken with greater accuracy than those in the original list, have been used. These are:

w_1. Maximum breadth outside condyles avoiding excrescences on these processes. This maximum projection may be taken in any direction and it is not necessarily horizontal or transverse.

$c_y l$. Maximum projective length of the left condyle avoiding excrescences on these processes. This may be taken in any direction.

rb'. Minimum antero-posterior breadth of the left ramus at any inclination to the horizontal, but with the posterior terminal never less than 13 mm. distant from the gonion.

$m_2 p_1$. Chord between the points on the outer left alveolar margin at the middle of the second molar (or its cavity) and at the middle of the first premolar (or its cavity).

h_1. Symphyseal height from intradental to the point farthest removed from it in the symphyseal plane, this plane being determined by anatomical appreciation.

zz. Minimum chord between the anterior margins of the right and left *foramina mentalia*.

$c_r c_r$. Coronial breadth from right coronion to left coronion. If both condyles are missing, the coronia (the tips of the coronoid processes) cannot be located with sufficient accuracy to justify the measurement being taken.

The above seven are caliper measurements and all those below, except the last, are taken with the aid of a mandible board of which photographs are given in the paper describing the technique.

$M\angle$. Mandibular angle, i.e. the angle between the standard horizontal and standard rameal planes.

$c_p l$. The projective length of the corpus.

rl. The projective length of the left ramus.

* "A Study of an Egyptian Series of Mandibles, with Special Reference to Mathematical Methods of Sexing," *Biometrika*, XXVIII (1936), pp. 149–78.

† Comparisons have not been made with the Anglo-Saxon series published by J. C. Brash, Doris Layard and Matthew Young (*ibid.* XXVII (1935), pp. 398–404) as the constants for it were not available at the time when the calculation for the present paper was carried out.

‡ The technique is described and full definitions of the measurements finally adopted are given in the Appendix of his 1936 paper cited.

ml. The maximum projective length of the mandible. Both condyles make contact with the vertical rameal wing of the board, and the solid set-square makes contact with the most advanced point of the chin.

$c_r h$. Projective height of the left coronoid process.

$m_2 h$. Projective height of the corpus at the middle point of the outer alveolar margin of the second left molar.

$R\angle$. Angle of condylar-coronoidal line with ramus tangent. If either the condylar or coronoid process is defective on the left side, then the mandible is positioned from the right side.

$g_o g_o$. Chord from left gonion to right gonion, found with small calipers, the mandible board being used to locate the gonia.

The last measurement is taken with a goniometer.

$C'\angle$. Mental angle, i.e. the angle between the standard horizontal plane and the line joining the infradental to the most anterior point in the standard sagittal plane of the symphysis (pogonion). The infradental is defined to be the mid-point of the common tangent to the two curves of the outer alveolar margins of the central incisors.

The question of personal equation in regard to the characters used in this paper was investigated by G. M. Morant in his paper "A Biometric Study of the Human Mandible".* The choice of characters recommended in that paper, chiefly on the ground that they had been shown to be the most reliable of all the characters originally defined in the technique, was accepted by the present writer who followed the amplified definitions given in the Appendix.

In order to obtain estimates of personal equation, one series—the Spitalfields —was measured on two occasions. The first set of readings (Cl_1) were those obtained by the writer when he was starting to measure mandibles, and the second set (Cl_2) comprise those readings taken by the writer after nearly three months' experience of the measurements in the Galton Laboratory. Part of the data in Table I is based on 50 comparisons of these two sets for each measurement. The measurements (H) taken on the Farringdon Street mandibles by Miss B. G. E. Hooke† were also available for comparison with the measurements (Cl), 16 in number, taken by the present writer on the same series. The remainder of the data in Table I is based on comparisons of these latter two sets for each measurement. Miss Hooke followed the original definitions given in *Biometrika*, XIV (1923), and, unlike the present writer, she did not work under the direct supervision of Dr G. M. Morant. Comparisons in connection with personal equation can further be made with the data for repeated measurements published by him‡. He deals with the personal equation involved in taking two series of

* *Biometrika*, XXVIII (1936).
† *Ibid.* XVIII (1926).
‡ *Loc. cit.* (1936), Table I.

measurements himself on the same material ($M_1 - M_2$), an interval of two years having elapsed between the times when the first and second measurements were taken, and that involved when his measurements are compared with those of Miss M. Collett on the same series ($M_1 - C$). The measurements of all the above series were taken on mixed series of male and female mandibles, as it is reasonable to assume that the personal equation is likely to be the same for both sexes.

TABLE I

Data for estimating the personal equation of mandibular measurements

Characters	Maximum individual differences		Differences of means (Δ)		Standard deviations of differences	
	$H-Cl$*	Cl_1-Cl_2†	$H-Cl$	Cl_1-Cl_2	$H-Cl$	Cl_1-Cl_2
w_1	+1·4	+1·1	+0·04 ± 073 (30)	−0·08 ± ·044 (50)	0·59 ± ·051	0·46 ± ·031
$g_o g_o$	+1·7	+&−1·8	+0·36 ± ·055 (51)	+0·13 ± ·073 (50)	0·58 ± ·039	0·77 ± ·052
$c_r c_r$	−1·3	—	+0·01 ± ·051 (36)	—	0·45 ± ·036	—
zz	+1·6	+&−0·9	−0·15 ± ·043 (50)	+0·07 ± ·030 (50)	0·45 ± ·030	0·31 ± ·021
$c_y l$	−1 1	+1·5	−0·21 ± ·045 (32)	−0·13 ± ·034 (50)	0·38 ± ·032	0·36 ± ·024
ml	+4·1	−1·5	+0·80 ± ·092 (47)	+0·03 ± ·063 (50)	0·94 ± ·065	0·66 ± ·045
$c_p l$	+3·3	+1·2	+1·04 ± ·094 (51)	+0·08 ± ·041 (50)	1·00 ± ·067	0·43 ± ·029
rb'	+2·0	−0·9	+0·29 ± ·054 (47)	+0·10 ± ·022 (50)	0·55 ± ·038	0·23 ± ·016
$m_2 p_1$	−2·2	+&−0·8	−0·07 ± ·095 (33)	−0·12 ± ·032 (50)	0·81 ± ·067	0·34 ± ·023
h_1	−3·3	—	−0·44 ± ·088 (29)	—	0·70 ± ·062	—
$m_2 h$	−1·2	+2·3	−0·27 ± ·075 (23)	−0·02 ± ·059 (50)	0·53 ± ·053	0·62 ± ·042
$c_r h$	−1·5	−1·7	−0·26 ± ·053 (46)	−0·06 ± ·054 (50)	0·53 ± ·037	0·57 ± ·038
rl	+2·7	+1·7	+0·75 ± ·077 (47)	+0·02 ± ·069 (50)	0·78 ± ·054	0·72 ± ·049
$M\angle$	+3°·5	−3°·0	−0°·16 ± ·110 (51)	−0°·60 ± ·081 (50)	1°·16 ± ·077	0°·85 ± ·057
$R\angle$	−3°·0	−2°·0	+0°·53 ± ·131 (39)	−0°·40 ± ·083 (50)	1°·21 ± ·092	0°·87 ± ·059
$C'\angle$	+6°·0	—	+1°·20 ± ·385 (28)	—	3°·02 ± ·272	—

* Differences for the Farringdon Street series. † Differences for the Spitalfields series.

The maximum differences found are given in columns 2 and 3 of Table I of the present paper, and in the same columns of Table I in Morant's paper. These maximum differences are of much the same order for the sets of differences ($M_1 - C$), ($M_1 - M_2$) and ($Cl_1 - Cl_2$), but for the set ($H - Cl$) they are considerably larger for the characters h_1, ml, $c_p l$, $m_2 p_1$ and $C' \angle$. Restricting comparisons to the 13 characters whose differences are available for all four sets, it is found that for seven of these ($H - Cl$) has the greatest maximum difference. The effect of personal equation on mean values may now be considered. Columns 4 and 5 of Table I of the present paper and of Table I of Morant's paper give the differences of the means (Δ) for the four sets of differences and the probable errors of these constants. In the set of differences ($M_1 - C$) 9 of the total 16 characters have values which differ from zero by less than three times their probable errors. In the set ($M_1 - M_2$) there are 11 out of the 16 showing the same relationship. In the set ($Cl_1 - Cl_2$) there are 8 out of 13 showing the same relationship. In the set

$(H-Cl)$ 4 characters only from a total of 16 have values which differ from zero by less than three times their probable errors. It is again evident that the differences $(H-Cl)$ are clearly distinguished from the other three sets. It will be sufficient in comparing mean values of the differences for the different sets to use the (M_1-C) values and to ignore the values of (M_1-M_2), as it has been shown by Morant that the former are on the whole the more reliable. In the comparison of Δ_{M_1-C} and $\Delta_{Cl_1-Cl_2}$ for the 13 characters possible, 2 of the differences, irrespective of signs, exceed three times their probable errors. These are for $c_r h$ (difference/p.e. difference $= 3{\cdot}5$) and $M\angle$ $(3{\cdot}8)$ In the first of these cases the difference (M_1-C) is the greater, but for $M\angle$ the reverse is true. Thus the only character for which the standard of accuracy of the present writer is appreciably less than that previously demanded—a demand made in consideration of Dr Morant's statistical treatment of the measurements dealt with in his paper—is $M\angle$. An explanation of this fact is found in the circumstances under which the measurement of this character took place. The Spitalfields series, on which the measurements were taken to obtain the set of differences (Cl_1-Cl_2), contains a large percentage of mandibles lacking a condylar process, or having one of these processes badly damaged. The positioning of the mandible for the measuring of $M\angle$ is a matter of approximation in these cases, and such approximation on the mandible board is quite likely to give rise to unusually large errors in taking the measurements. Now, after subsequent laboratory experience, the present writer would not include measurements as doubtful as some of those recorded for the Spitalfields series, and therefore it seems reasonable to suggest that the inaccuracies consequent upon taking too many doubtful measurements are responsible for the unsatisfactory nature of his earlier readings of $M\angle$. Comparisons irrespective of signs may now be made between Δ_{M_1-C} and Δ_{H-Cl}. Out of the 16 comparisons possible 6 differences are greater than $3{\cdot}5$ times their probable errors. These are for $c_y l$ $(3{\cdot}9)$, $g_o g_o$ $(4{\cdot}3)$, $c_r c_r$ $(5{\cdot}3)$, ml $(6{\cdot}8)$, rl $(7{\cdot}6)$ and $c_p l$ $(8{\cdot}0)$. In all except the third of these cases Δ_{H-Cl} is greater than Δ_{M_1-C}. Three of these measurements are taken on the mandible board, and it is clear that Miss Hooke had a conception of the definitions different from that of the present writer, all her measurements tending to be greater than his.* A detailed comparison of the differences of the means (Cl_1-Cl_2) and $(H-Cl)$

* That this difference in interpretation of the definitions exists in the case of these two workers is further illustrated by an examination of the numbers of mandibles on which either took any one of the 3 significantly different mandible board measurements, while the other omitted it. Such an examination shows that Miss Hooke took 23 measurements of these characters where the present writer did not, and that the latter took 11 of the bilateral measurements (rl) on the right where Miss Hooke took them on the left. It is clear that all estimates of personal equation are likely to be considerably influenced by measurements taken on imperfect specimens, and that variabilities of differences will be very much reduced if questionable readings are omitted. In the present instance wider discrepancies would have been evident in the set of differences $(H-Cl)$, but for the fact that throughout the measurement of the whole series the present writer refrained from taking measurements in 104 doubtful cases for which Miss Hooke had given readings.

need not be made, for in 10 cases out of 13 the latter are the greater. It is thus again evident that the differences $(H - Cl)$ are of far greater account than any other differences available.

Comparing the standard deviations of the differences in the same way, it is found that for 4 of the 13 characters the differences $(M_1 - C)$ and $(Cl_1 - Cl_2)$ exceed 3·5 times their probable errors. These are for m_2h (3·7), g_0g_0 (3·8), rb' (5·6) and c_yl (5·9). In all cases except the third the standard deviation $(Cl_1 - Cl_2)$ is in excess. A comparison of the standard deviations of the differences $(M_1 - C)$ and $(H - Cl)$ shows that 7 are significant from a total of 16 characters. These are for c_yl (5·1), m_2p_1 (5·2), $R\angle$ (5·4), $C'\angle$ (5·5), zz (5·9), c_pl (7·3) and w_1 (7·4). In all these cases the standard deviation $(H - Cl)$ is the greater. A detailed comparison of the standard deviations of the differences for $(Cl_1 - Cl_2)$ and $(H - Cl)$ need not be made for, just as in the case of the differences of the means, in 10 comparisons out of 13 the $(H - Cl)$ constant is greater than the corresponding $(Cl_1 - Cl_2)$ constant.

It is evident, from the comparisons made between mean differences and standard deviations of differences, that the two sets of readings taken on the same English mandibles by the present writer indicate errors of personal equation which are almost precisely of the same order as those found between the readings taken by Miss Collett and G. M. Morant on an Egyptian series. Where the most significant differences between the corresponding constants were found —viz. in the case of a few of the standard deviations of the differences—the readings $(Cl_1 - Cl_2)$ are slightly less consistent than the readings $(M_1 - C)$. It was shown by Morant in his paper that the errors indicated in the latter case $(M_1 - C)$ for the characters used in the present paper are not large enough to invalidate inter-racial comparisons, and it seems safe to assume that the same will be true for the readings taken by the present writer.*

* The measurements used in this paper were selected by Morant from a larger number originally defined mainly on the grounds that they were found to be the most accurate ones. The tests used in making the selection depended on comparison of the differences of means found between two sets of readings on the same mandibles with the probable errors of the means of an Egyptian series, and on a second comparison of the standard deviations of the differences with the standard deviations of the same Egyptian series. Full details of the method used are given in his paper. Applying the same tests to the set of differences $(Cl_1 - Cl_2)$, it is found that 4 out of 13 characters fall short of the standard accepted in that paper. These are c_yl, m_2h, $M\angle$ and m_2p_1. The last did not satisfy the tests in the case of Morant's own data, but its continued use was recommended, since it is but little less reliable than the other characters accepted and it is a measurement of particular interest. The lack of reliability in the case of the characters c_yl and m_2h may perhaps be explained by the failure on the part of the present writer to reject the specimens on which it was doubtful whether a close enough approximation of the measurement could be obtained, and by his inability to deal effectively with the condylar anomalies met with in taking the measurement c_yl on the Spitalfields mandibles—the first series he measured. In the case of the differences $(H - Cl)$ 10 of the 16 characters fail to fulfil the requirements of the tests, and for this comparison the measurement $C'\angle$ is found to be the least reliable. It is, therefore, of interest to note that for the differences between Dr Morant's and the author's readings for this angle the tests are satisfied, although they were not satisfied for Morant's and Collett's original data (see Table II below).

When my readings on the Farringdon Street mandibles are compared with those of Miss Hooke on the same series, differences of a markedly higher order are found, whether the means or the standard deviations of the differences are considered. The same discordance is found when the set of differences $(H-Cl)$ is compared with the set (M_1-C). It is, therefore, reasonable to suppose that Miss Hooke's measurements on any series, when compared with those of any one of the other three measurers on the same series, would not have agreed with theirs as closely as theirs would have with one another. It is not possible to assert that she measured less *accurately* than they did, since the relations observed above between her measurements and those of the other three measurers are probably due to the fact that she had only the original unrevised definitions as a guide to her measuring, and that she was not able to work in consultation with anyone who had previously applied the technique.

To close this section on personal equation, a comparison is made between the measurement of $C'\angle$ (hitherto regarded as the most unsatisfactory of the characters included in the technique used in the present paper), taken by Morant and by the writer. This measurement was recorded for three series of mandibles by both these workers on account of its suspected unreliability. The results are set out in Table II below. The mean difference for the combined series differs significantly from zero, but both it and the standard deviation of the differences are less than the corresponding constants (M_1-C), though not significantly so.

TABLE II

Data for estimating the personal equation of the mental angle $(C'\angle)$

Series	Maximum individual differences	Differences of means (Δ)	Standard deviations of differences
	$M-Cl$	$M-Cl$	$M-Cl$
Australian	+ & − 2°·5	+0°·51 ± 096 (62)	1°·12 ± ·068
Punjabi	+2°·5	+0°·41 ± ·138 (28)	1°·08 ± ·097
Farringdon Street	+3°·0	+0°·29 ± ·114 (43)	1°·11 ± ·081
Combined	+3°·0	+0°·42 ± ·065 (133)	1°·11 ± ·046

On comparison with the constants of the set of differences $(H-Cl)$ in Table I, for the character $C'\angle$, it is found that the mean for the combined differences $(M-Cl)$ is the smaller, though not significantly so, while the standard deviation of the differences for $(M-Cl)$ is also the smaller, and markedly so, the difference for $(H-Cl)$ and $(M-Cl)$ in this case being 6·9 times its probable error. From consideration of the above results, it seems reasonable to assume that a marked improvement has occurred in the accuracy with which the mental angle has been

taken. Apparently the difficulties attending the measurement of this important character can be overcome by care and practice in measurement, and further improvement in accuracy would be expected if better designed instruments were substituted for those now in use, which are far from satisfactory.

4. *Methods of sexing the material.* In cases where the sexes are not known, it is probably impossible to secure absolute accuracy in sexing a series of crania or mandibles by anatomical inspection or any other method. Assessment of the sex depends on the nature of the whole series dealt with, and not on standardized conceptions of characters of the bone that remain constant for every possible series. The sexing of any series of bones available is, however, of great importance, as the statistical treatment of osteometric material demands that each sex be considered separately. Mathematical methods of sexing have therefore been devised to supplement anatomical sexing, such methods being based on the combined values of certain metrical characters of the bones. It can be assumed that the distribution of any of these particular characters for either sex, in the case of a homogeneous series, will be approximately normal. The most suitable characters for discriminatory purposes are those whose means differ most in proportion to the standard deviations of the distributions for the two sexes. The characters chosen should also have low intra-racial correlations, and it is an advantage if they can be found for a high percentage of bones, so that as few specimens as possible will have to be left unsexed.

Dr E. S. Martin has discussed several methods of sexing mandibles in a recently published paper.* He came to the conclusion that the most effective characters for sexing purposes, and those which fulfilled the above conditions most adequately, were $g_o g_o$, $c_p l$, $c_r h$ and $M\angle$, and these have been used for the purpose in the present paper. He also showed that anatomical sexing is far more reliable than had been generally supposed, the percentage agreement between mathematical and anatomical sexing being so high that considerable reliance can be placed on anatomical sexing alone. The method finally used in sexing two of the series of mandibles dealt with in this paper was a combination of a mathematical method with that of anatomical inspection in the cases where the sex was doubtful. Dr Martin demonstrated that very little difference is made to the accuracy of mathematical sexing by the inclusion of aged mandibles in a series, and hence no account has been taken of the relative ages of the adult mandibles. The sexing of the mandibles dealt with in this paper was carried out before Dr Martin's work was published, and hence the more elaborate methods of sexing given in his paper were not applied. Moreover, the series dealt with here are so short that it seems reasonable to suggest that there would be very little practical advantage in applying the more elaborate methods which he found to give only slightly better results. It will, however, be interesting to note how far the

* *Loc. cit.*

admittedly cruder mathematical methods here employed give results in agreement with anatomical sexing.

For the purpose of sexing the mandibles, it was assumed that the proportion of males to females in the sample to be sexed is the same as that of the cranial series to which it belongs. The less crude of the two mathematical methods of sexing used—referred to here as method I—is concerned, theoretically, with the sum of the four ratios obtained by dividing the deviations from the means of the characters, $g_o g_o$, $c_p l$, $c_r h$ and $M\angle$, by the corresponding standard deviations for the total series. The mean values for the characters $g_o g_o$, $c_p l$ and $c_r h$ are higher for males than females, but the reverse is true for $M\angle$, and hence the ratio for this last character has to be subtracted from the sum of the ratios of the other three for each mandible. There are 95 Spitalfields mandibles, and if the proportions of the sexes are to be supposed the same for these as for the crania we must take 32 as female and 63 as male.* The 32 with the lowest scores will be counted female. In actual practice, however, the absolute measurements themselves, and not their deviations from the means, were divided by the standard deviations for the total series, since the mandibles are arranged in the same order by these two procedures and the former entails less calculation.

The second method tried proceeds as follows: the measurements of each mandible for the characters $g_o g_o$, $c_p l$, $c_r h$ were added together, and in each instance the measurement of $M\angle$ was subtracted from this total despite the difference of units of measurement used, viz. millimetres for the first three and degrees for the last character. The 32 mandibles with the lowest totals were classed as female. This method will be referred to as method II.

The Spitalfields series was sexed anatomically by Dr G. M. Morant, who in doing this accepted the proportions of males to females given by the crania, and it will be interesting to note the percentage agreements between the methods used. These are 87·4 between inspection and method I; 85·3 between inspection and method II; and 95·8 between method I and method II. The present writer also sexed the Spitalfields series anatomically, and there is an agreement between this sexing and that obtained from the application of method I in 83·2 per cent. of cases. It is surprising to find that such a high percentage agreement between anatomical and mathematical sexing is obtained in the case of method I. This is a crude method, since the four characters used are assumed to be of equal importance for sexing purposes and no account is taken of correlations between them. It is more surprising still to find a percentage agreement nearly as high when method II is used, since this method is wholly unsatisfactory from a theoretical point of view.† As the above percentage agreements between anatomical and

* Of the 883 adult Spitalfields crania which were sexed, 590 (66·8 per cent.) were supposed male and 293 (33·2 per cent.) female.

† If either $c_r h$ or $c_p l$ is taken singly as a criterion of sex there is agreement between these estimates based on single characters and anatomical sexing of 76·8 per cent. in both cases, and an agreement between the two of 66·3 per cent.

metrical estimates are of the same order as those obtained from the more elaborate methods discussed by Dr Martin, it seems fairly reasonable to conclude that the crude mathematical method (I) is satisfactory. The mandibles of the Spitalfields series which had been classed oppositely by method I and that of anatomical inspection were re-examined anatomically, and on this re-examination the sex of the doubtful cases was decided. The crania to which 12 of the mandibles of the Spitalfields series had been attached were known, and these had been sexed anatomically. The accepted sexes of these 12 mandibles (all those numbered under 1000) were now compared with the results obtained from an anatomical sexing of the corresponding skulls, and in 9 cases agreement was found. Three cases disagreed, viz. No. 509 sexed as male mandible—skull sexed female; No. 401 sexed female mandible—skull sexed male; No. 507 sexed female mandible—skull sexed male. In each of these cases both methods of sexing the mandible gave the same result, which was accepted in spite of the disagreement with the sexing of the cranium.

The original population from which the Farringdon Street series, comprising 90 mandibles, was drawn is represented by 381 crania sexed in the ratio 213 female and 168 male. Using the same ratio we thus obtain for the mandibles 50 female and 40 male. The percentage agreement obtained between the mathematical method (I) of sexing described and that of anatomical inspection by Dr G. M. Morant was 88·9 per cent., and by the writer 86·7. There was an agreement of 86·7 per cent. between the two anatomical estimates. A re-examination of the 10 doubtful cases anatomically, as in the case of the Spitalfields series, finally decided their sexes.

A percentage agreement between anatomical and mathematical methods of sexing lies at best somewhere between 85 and 90 per cent., and it is hardly possible to improve upon this on account of the presence of border-line cases in the samples.* The most satisfactory way to sex a series of mandibles, for which the ratio of males to females is assumed known, seems to be that of re-examination of those mandibles which have been sexed oppositely by the two methods, and finally sexing these cases mainly on anatomical grounds. The mathematical method is then merely a subsidiary one used to support anatomical sexing. It should be noted that the method of sexing adopted in the case of the two English series depends essentially on the assumption that the proportions of the sexes are the same for the mandibular as for the corresponding cranial samples. If this assumption is incorrect some of the bones will inevitably be sexed incorrectly.

An examination of the sex ratios of the absolute measurements (i.e. male mean/female mean) for the Spitalfields, Farringdon Street and Australian series shows that for any particular ...

ratios for the different series. The small numbers that go to make up most of the female series, however, give very little weight to any conclusions drawn from such ratios. Nevertheless, it is interesting to note that the generally accepted idea that the differences of sex are more marked among primitive than among civilized peoples is not borne out by the ratios for the three series concerned, the Australian ratio being the greatest only in 4 (and equal to the Farringdon Street in 2) out of 13 comparisons. The heights have the largest ratios, but they are not markedly higher than those for the remaining measurements. The sex ratios tend to be higher for mandibular than for cranial characters in the case of a particular series. This has been shown in the case of the Kerma from Egypt, and it is also true for the Spitalfields and Farringdon Street series.

5. *Racial differences in variability.* It is possible to draw an unambiguous conclusion as to relative racial variability with respect to a measurable character of one series compared with the corresponding character of another, and one approach to the problem of variability would be to make no statement concerning relative racial variability, unless it referred merely to variabilities of single characters in different series. An alternative method of approach is to assume that an estimate of general variability for all characters measured can be obtained. When, however, a large number of characters are considered, any statement about relative total variability (i.e. variability for all characters considered together) must be somewhat arbitrary, since it depends on the particular method of comparison employed. When one race shows a consistently higher variability than another, in the case of all characters showing a significant difference, we can reasonably assume that it is the more variable. When an equal number of significantly different characters are found for each race in excess of the other, then it seems impossible to assign greater variability to either.

In Tables III and IV are given the standard deviations for all characters and the coefficients of variation for the absolute measurements, respectively. Using these coefficients for absolute measurements and the standard deviations for angles and indices, a comparison for characters considered singly may be made between the four series of mandibles for which constants are given in the tables. The results of such comparisons for the four new series and based on the 21 characters are set out in Table V. In column 2 is given, for each series in a particular comparison, the number of characters having the greater constants of variability. In column 3 are shown the characters which differ significantly in the comparison, and, where such a significant difference occurs, the series for which the variability of the significant character is the greater. An examination of

TABLE III

*Standard deviations for series of mandibles**

Character	Standard deviations						
	Male				Female		
	Spitalfields	Farringdon Street	Punjabi	Australian	Spitalfields	Farringdon Street	Australian
w_1	5·40±·49	3·75±·37	5·77±·40	5·97±·39	3·77±·42	5·16±·45	6 28±·59
g_og_o	6 87±·41	6·58±·50	5·21±·36	7·83±·49	5·14±·43	5·74±·39	5·61±·56
c_rc_r	5·38±·38	5·44±·48	4·56±·32	5·97±·39	3·56±·35	4·75±·38	6·30±·57
zz	2·43±·15	2·21±·17	2·19±·15	2·49±·15	2·24±·19	2·56±·17	2·79±·22
c_yl	1·68±·14	1·72±·14	1·83±·12	2·01±·12	1·32±·13	1·55±·11	2·29±·20
ml	5·41±·40	5·51±·45	6·27±·44	3·92±·24	4·50±·41	5·97±·43	4·88±·43
c_pl	3·98±·24	3·76±·28	4·54±·32	5 05±·31	3·10±·26	3·93±·27	4·55±·39
rb'	2·37±·14	2·60±·20	2·59±·18	3·01±·19	1·98±·17	2·61±·18	2·79±·23
m_2p_1	1·94±·13	1·22±·12	1·63±·13	1·52±·09	1·06±·12	1·44±·16	1·42±·12
h_1	2 49±·18	2·34±·32	2·15±·24	2·97± 20	3·07±·32	2·73±·26	2·36±·23
m_2h	1·68±·11	2·88±·32	2·72±·23	2·42±·15	2·29±·24	2·60±·36	2·43±·22
c_rh	5·05±·30	4·38±·33	5·20±·36	5·13±·32	3·86±·33	4·91±·33	4·93± 40
rl	5·02±·33	3·54±·28	4·22±·29	5·10±·32	3·26±·29	4·23±·31	5·39±·46
$M\angle$	7°·03± 42	5°·67±·43	5°·65±·39	6°·55±·41	5°·84±·49	6°·21±·42	5°·01±·43
$R\angle$	7°·74±·53	8°·50±·67	7°·88±·54	7°·10±·44	6°·08±·56	8°·96±·62	5°·69± 50
$C'\angle$	6°·67±·47	7°·99±·95	8°·36±·85	5°·68±·43	6°·73±·70	6°·13±·56	5°·16±·55
$100\ c_rh/ml$	4·77±·35	5·73±·47	6·16±·43	5·47±·35	3·94±·36	4·79±·35	4·73±·42
$100\ c_rc_r/ml$	6·27±·51	7·67±·73	6·02±·43	6·69±·44	4·98±·50	6·03±·53	7·14±·67
$100\ g_og_o/c_pl$	11·59±·70	11·55±·87	10·37±·72	11·92±·77	10·98±·93	10·72±·72	9·70±·96
$100\ rb'/rl$	5·14±·34	4·82±·38	7·01±·49	4·99±·31	4·06±·37	6·45±·47	6·23±·53
$100\ g_og_o/c_rc_r$	7·25±·51	8·57±·76	5·51±·39	8·51±·59	6·33±·62	7·53±·61	8·19±·83

* The numbers of mandibles on which the constants in this table and in Table IV are based can be seen from the table of means (Table IX).

TABLE IV

Coefficients of variation for series of mandibles

Character	Coefficients of variation						
	Male				Female		
	Spitalfields	Farringdon Street	Punjabi	Australian	Spitalfields	Farringdon Street	Australian
w_1	4·50±·41	3·19±0·32	4·96±·35	4·04±·27	3·33±0·37	4·73±0·42	5·65±0·53
g_og_o	7·13±·43	6·73±0·51	5·61±·39	8·21±·52	5·86±0·49	6·70±0·45	6·47±0·64
c_rc_r	5·57±·39	5·67±0·51	4·84±·34	6·30±·41	3·92±0·38	5·18±0·42	7·17±0·65
zz	5·39±·33	5·03±0·38	4·99±·34	5·25±·30	5·16±0·44	5·94±0·40	6·09±0·49
c_yl	8·19±·67	8·69±0·72	8·97±·61	9·39±·59	6·95±0 68	8·61±0·62	11·80±1 03
ml	5·28±·39	5·29±0·43	6·12±·43	3·63±·22	4·57±0 41	6·01±0·44	4·77±0·42
c_pl	5·37±·32	4·93±0·37	6·10±·43	6·07±·38	4·58±0·39	5·62±0·38	5·99±0·52
rb'	7·36±·44	8·41±0·63	8·41±·58	8·77±·52	6·87±0·58	9·22±0·62	8·83±0·73
m_2p_1	6·93±·49	4·33±0·44	5·70±·47	5·00±·30	4·03±0·44	5·16±0·56	4·86±0·42
h_1	7·71±·55	7·57±1·05	6·44±·72	8·92±·61	10·62±1·12	9·19±0·86	7·66±0·74
m_2h	6·40±·44	11·57±1·28	10·54±·88	9·20±·57	9·79±1·03	11·02±1·54	9·92±0·90
c_rh	7·79±·47	6·75±0·51	7·90±·56	8·02±·50	6·75±0·57	8·67±0·58	8·82±0·72
rl	8·04±·53	5·69±0·45	6·75±·47	8·11±·51	5·72±0·52	7·91±0·58	9·57±0·83

the Farringdon Street and the Australian, since the former shows the lesser variability for most of the characters in these comparisons.

A comparison (see Table VI) was next made for another group of four series, viz. two Egyptian (the Egyptian *E* and Kerma, the latter having been stated to be slightly more variable than another Egyptian series from Qau*), and the Australian and Spitalfields, which are assumed on such evidence as is afforded in Table V to be the most and the least variable, respectively, of the series dealt with

TABLE V

Comparisons of the variabilities of two English, an Indian, and an Australian series of mandibles†

Series		Nos. of greater constants	Significant differences
Male:			
Spitalfields (SF.)	: Farringdon Street (FA.)	SF. 11 >, FA. 10 >	m_2p_1 (3·9) SF. >, m_2h (3·8) FA. >
Spitalfields	: Punjabi (Pu.)	SF. 10 >, Pu. 11 >	m_2h (4·2) Pu. >
Spitalfields	: Australian (Aus.)	SF. 8 >, Aus. 13 >	ml (3·7) SF. >, m_2h (3·8) Aus. >
Farringdon Street	: Punjabi	FA. 10 >, Pu. 10 >, 1 =	w_1 (3·7) Pu. >, 100 rb'/rl (3·5) Pu. >, 100 g_og_o/c_rc_r (3·6) FA. >
Farringdon Street	: Australian	FA. 7 >, Aus. 14 >	rl (3·6) Aus. >
Punjabi	: Australian	Pu. 9 >, Aus. 12 >	g_og_o (4·0) Aus. >, ml (5·2) Pu. >, 100 g_og_o/c_rc_r (3·9) Aus. >
Female:			
Spitalfields	: Farringdon Street	SF. 3 >, FA 18 >	100 rb'/rl (3·9) FA. >
Spitalfields	: Australian	SF. 5 >, Aus. 16 >	w_1 (3·6) Aus. >, c_rc_r (4·3) Aus. >, rl (3·9) Aus. >, c_yl (3·9) Aus. >
Farringdon Street	: Australian	FA. 12 >, Aus. 9 >	$R\angle$ (4·1) FA. >

† The constants compared are coefficients of variation of the absolute measurements (Table IV) and standard deviations of the indices and angles (Table III).

there. There seems to be no marked difference between the Kerma and the Australian series, though the Kerma, it is interesting to note, appears to be considerably more variable than the other Egyptian series (Egyptian *E*), as does also the Australian, especially in the comparison for the female series. A comparison of the two series which appear to show the least variability (viz. the

* G. M. Morant, *loc. cit.* p. 102.

TABLE VI

Comparisons of the variabilities of two Egyptian, an Australian, and an English series of mandibles

Series	Nos. of greater constants	Significant differences
Male:		
Kerma (K): Australian (Aus.)	K 10>, Aus. 11>	—
Kerma : Egyptian *E* (Eg.)	K 16>, Eg. 5>	rl (4·2) K.>
Kerma : Spitalfields (SF.)	K. 15>, SF. 6>	—
Egyptian *E*: Australian	Eg. 6>, Aus. 14>, 1=	c_yl (4·6) Aus.>, ml (4·1) Eg.>
Egyptian *E*: Spitalfields	Eg. 8>, SF. 12>, 1=	m_2h (4·6) Eg.>
Spitalfields : Australian	SF. 8>, Aus. 13>	ml (3·6) SF.>, m_2h (3·8) Aus.>
Female:		
Kerma : Australian	K. 11>, Aus. 10>	$M\angle$ (3·7) K.>
Kerma : Egyptian *E*	K. 17>, Eg. 3>, 1=	$100\,c_rh/ml$ (3·6) K.>, c_rh (5·0) K.>
Egyptian *E*: Australian	Eg. 7>, Aus. 14>	c_rh (3·7) Aus.>, c_yl (3·8) Aus.>
Egyptian *E*: Spitalfields	Eg. 14>, SF. 7>	w_1 (4·3) Eg.>, c_rc_r (3·7) Eg.>
Spitalfields : Australian	SF 5>, Aus. 16>	w_1 (3·6) Aus.>, c_rc_r (4·3) Aus.>, c_yl (3·9) Aus.>, rl (3·9) Aus.>

Spitalfields and the Egyptian *E*) indicates that for the female series the Egyptian *E* may be considered the more variable.

Although there appear to be racial differences in variability, the material available for the mandible accords with the far more extensive material relating to cranial and living series, in showing that the absolute differences between the variabilities of different races are exceedingly small. It is interesting to observe from Tables V and VI that the Australian and Egyptian tend to be more variable than the two English series. This is an unexpected result, but it may be due to some peculiar selection of the mandibles preserved.

6. *Special topics: correlations, asymmetry and records relating to teeth.* The coefficients of correlation between the various mandibular measurements throw some light on the interdependence during growth of various parts of the mandible.* The first structural peculiarity we note from an examination of the correlations in Morant's paper is the fact that the dental arcade between the mid-points of the alveolar margins of the first premolar and second molar (m_2p_1) seems to be uncorrelated with any other chords. Apparently, growth of this particular area

* This section is complementary to that on correlation in G. M. Morant's paper, "A Biometric Study of the Human Mandible", *loc. cit.* pp. 103–8.

ceases at an early age. Thus the coefficients of correlation between m_2p_1 and all the antero-posterior chords were previously found to be insignificant: in the Australian male series, however, a significant value is found for m_2p_1 and c_pl ($r = +\cdot 328 \pm \cdot 081$). This would certainly have been expected *a priori* as c_pl "covers" m_2p_1.

Since the correlation coefficients for like measurements (i.e. breadths with breadths, heights with heights, etc.) are most interesting when they are insignificant, and since the converse holds good for unlike measurements, it is of interest to note (for the Qau Egyptian series) that though h_1 is, as expected, correlated with m_2h, yet it is not so with rl or with any other rameal height. The measurement (h_1) is, however, fairly highly correlated with ml, and this seems to point to the fact that an antero-posterior growth of the mandible necessitates a corresponding growth in height of the corpus. It is surprising to find a low correlation coefficient between c_pl and ml, two antero-posterior chords, especially as the former is "covered" by the latter. An unexpectedly high correlation is that between $R\angle$ and $100\, c_rc_r/ml$, though what growth factor influences these two particular measurements is not evident. Another result which reveals an unexpected feature of the architecture of the mandible is the high negative correlation found between the breadth of the ramus (rb') and the mandibular angle ($M\angle$). The broad solid ramus is apparently found on the upright looking mandible, while the slender ramus accompanies the sloping type of mandible.

The few correlation coefficients computed for the Australian male series by the writer, and set out below, lead to the same conclusions as those derived from the Qau series. A comparison between corresponding values for the two series reveals no single significant difference. The following coefficients are found for the Australian bones:

h_1 and $M\angle$ $+\cdot 090 \pm \cdot 102$ (43); h_1 and ml $+\cdot 516 \pm \cdot 075$ (44),
h_1 and m_2h $+\cdot 539 \pm \cdot 071$ (45), h_1 and m_2p_1 $+\cdot 015 \pm \cdot 097$ (48),
h_1 and rl $+\cdot 291 \pm \cdot 094$ (43); rb' and $M\angle$ $-\cdot 509 \pm \cdot 065$ (59),
$C'\angle$ and $M\angle$ $-\cdot 299 \pm \cdot 100$ (38); m_2h and m_2p_1 $-\cdot 064 \pm \cdot 088$ (58),
c_pl and ml $+\cdot 390 \pm \cdot 076$ (57); c_pl and m_2p_1 $+\cdot 328 \pm \cdot 081$ (55),
$R\angle$ and $100\, c_rc_r/ml$ $+\cdot 548 \pm \cdot 066$ (51).

It appears that all intra-racial correlations between absolute measurements of the mandible are positive, whether significant or not, or negative and insignificant. In other words, a large mandible tends to be large in all respects, and a small one to be small in all respects, a fact for which the normal growth of the mandible as a whole is evidently responsible. The measurement m_2p_1 alone shows no tendency to conform to the general growth trend.

The asymmetry of the mandible has sometimes been taken as a well-established fact, and anatomists as famous as Le Double have even made categorical statements to the effect that the right side of the mandible is on the average always

greater than the left. An examination of the bilateral measurements—viz. m_2p_1, $c_y l$ and rb'*—which were taken on the mandibles of the Qau and Kerma male and female series has shown that there is fairly clear evidence of a slight asymmetry in type for these Egyptian series, and that the right side of the mandible is not consistently greater than the left. In fact, the length of part of the dental arcade (m_2p_1) as well as the breadth of the ramus (rb') were larger on the left side than on the right, the length of the condyle ($c_y l$) alone supporting Le Double's hypothesis. It is strange to find that the broader ramus supports the smaller condyle (at least if the length of the condyle is any criterion of its size). From Table VII it will be seen that the statistical evidence warrants no assumption of asymmetry in type in the case of the Australian mandible. The means actually found are all slightly greater on the right than on the left, but the bilateral differences are quite insignificant.

TABLE VII

Constants of bilateral differences for the Australian male series

Means ($L-R$)			Standard deviations		
m_2p_1	$c_y l$	rb'	m_2p_1	$c_y l$	rb'
$-\cdot079 \pm \cdot075$ (57)	$-\cdot028 \pm \cdot096$ (50)	$-\cdot006 \pm \cdot098$ (64)	$0\cdot84 \pm \cdot053$	$1\cdot01 \pm \cdot068$	$1\cdot16 \pm \cdot069$

Comparing the male Australian differences with those previously given for the two Egyptian series,† no clearly significant differences are found, so there is no evidence to show that there are racial differences in asymmetry.

Table VIII shows that a high percentage of the Australian mandibles, both male and female, had never lost a single tooth during life. This is in marked contrast to the civilized English series represented by the Farringdon Street (composed of seventeenth-century Londoners) and Spitalfields series (probably a population living in England during the Romano-British period). The Indian series has about 50 per cent. of its mandibles—25 out of 49—with a complete set of teeth, and, while being far below the corresponding percentage for the Australian series (approximately 80 per cent.), such a figure is much higher than that for the English series, which for males and females combined gives a percentage of 22·5.‡ There were several cases in the English and Indian series of arthritic condyles, though not a single instance of this was found in the Australian series which contains a case of syphilis (See Plate V A) according to the catalogue of the Royal College of Surgeons. The small numbers found in category 3 of Table VIII

* These measurements are usually taken, when possible, on the left side only.

† G. M. Morant, *loc. cit.* Table II.

‡ In the case of the long Egyptian *E* series examined by Dr Martin the percentages having all teeth, including third molars, present at death are 40·7 for males and 44·3 for females.

do not in any way affect the conclusions stated above and based on the figures found in category 2. It is to be noted that the third category of the table includes some cases for which the third molars had probably never erupted; for an examination of the dental arcade often fails to determine whether a molar was lost before death, or whether it had never erupted at all. The following cases of overcrowding were noted: Spitalfields male 3, female 2; Australian male 4, female 0; Farringdon Street male 1, female 1; Punjabi male 0. It appears that the opinion expressed by several anthropologists to the effect that overcrowding of the teeth is on the increase among modern civilized peoples is not supported by these figures, which

TABLE VIII

Comparisons of the dentitions of series of mandibles

	Male				Female		
	Spitalfields	Farringdon Street	Australian	Punjabi	Spitalfields	Farringdon Street	Australian
1. No. for which dental arcade is complete	27	69*	72†	49‡	22	83§	36
2. All teeth including 3rd molars present at death	10	18	58	25	5	12	26
3. All teeth except one or both 3rd molars present at death	5	8	3	1	4	8	2
4. No. having lost one or more teeth in front of molars before death	5	24	4	16	6	37	7

* 40 measured, 29 not measured; these latter were sexed anatomically.
† Including Northern Territory mandibles and 1 with socket for single pair of incisors.
‡ Including 2 with sockets for three incisors only.
§ 50 measured, 33 not measured; these latter were sexed anatomically.

refer only to the more marked cases.|| Furthermore, to assert on the alleged evidence of the increased incidence of overcrowding of the teeth that there is a tendency for modern man, civilized and uncivilized, to have a dental arcade smaller in size than that of his primitive ancestors would be unjustifiable, since overcrowding is dependent on the ratio of size of teeth to size of jaw, and not on the absolute size of the dental arcade.

7. *Comparisons by the method of the coefficient of racial likeness.* One of the primary needs of physical anthropologists in dealing with problems of human evolution is a means of classifying the races of man. To meet this need Professor Karl Pearson devised the method of the coefficient of racial likeness, which is a

|| Several cases of impaction of the third molar were also noted.

generalized statistical criterion derived from pairs of racial series and based on the comparison of a number of their mean measurements for different characters. Large numbers of these coefficients computed for different groups of cranial series have been published, chiefly in *Biometrika*, and the method has also been applied to data for series of living people in several papers. In his study of the mandible (*loc. cit.* 1936) Morant gives values for all possible pairs of 12 male and 5 female series, and he says (p. 116) that: "these criteria lead to a reasonable arrangement of the types which encourages the hope that a similar comparison of more extended material would furnish valuable aid in estimating racial relationships. At the same time the fact that samples of the sizes at present available are not differentiated cannot be accepted as a test of racial identity." The coefficients given by Martin (*loc. cit.* Table V) between his series of 26th–30th Dynasty mandibles from Gizeh and the earlier material do not conflict with these conclusions. We are now able to add comparisons with the 4 male and 3 female series described in the present paper, making a total of 17 male and 9 female, and it will be seen that these make it necessary to reconsider the position and, indeed, to question whether it is possible to obtain any rational classification of races from measurements of mandibles. The mean measurements of the new series are given in Table IX.

If M_s is the mean and σ_s the standard deviation of the sth character, these being based on n_s individuals, in the case of the first series, and if $M_{s'}$, $\sigma_{s'}$ and $n_{s'}$ are the corresponding constants for the second series, then what is now called the "crude" coefficient of racial likeness is defined to be:

$$\frac{1}{m}\Sigma\left\{\frac{(M_s-M_{s'})^2}{\dfrac{\sigma_s^2}{n_s}+\dfrac{\sigma_{s'}^2}{n_{s'}}}\right\}-1\pm\cdot67449\sqrt{\frac{2}{m}},$$

where m characters are compared. The standard deviations for the shorter series are likely to be particularly unreliable, and hence it is assumed that they are equal to those for the longest homogeneous series available. It has been shown above that, though these constants for different series show a few significant differences, yet they tend to be of closely similar orders in the case of a particular character. Supposing that $\sigma_s = \sigma_{s'}$, the coefficient becomes:

$$\frac{1}{m}\Sigma\left\{\frac{(M_s-M_{s'})^2}{\sigma_s^2}\times\frac{n_s n_{s'}}{n_s+n_{s'}}\right\}-1\pm\cdot67449\sqrt{\frac{2}{m}},$$

which is written, for convenience, as:

$$\frac{1}{m}\Sigma(\alpha)-1\pm\cdot67449\sqrt{\frac{2}{m}}.$$

Following Morant, the standard deviations of the ancient Egyptian series from Qau were used.* From the crude coefficient we can obtain a generalized measure

* The Qau is the longest of the series used in his paper. That of the 26th–30th Dynasty mandibles from a cemetery at Gizeh, described by Martin, is considerably longer, but the standard deviations for it were not available when the computation for the present paper was carried out.

TABLE IX

Mean measurements of series of mandibles

Character	Male				Female		
	Spitalfields	Farringdon Street	Punjabi	Australian	Spitalfields	Farringdon Street	Australian
w_1	120·0±·69 (28)	117·7±0·53 (23)	116·4±0·57 (47)	118·5±0·55 (53)	113·3±0·60 (18)	110·8±0·64 (30)	111·2±0·83 (26)
g_og_o	96·3±·58 (63)	97·7±0·70 (40)	92 8±0·51 (47)	95·4±0·70 (57)	87·7±0·61 (32)	85 7±0·55 (50)	86·7±0·79 (23)
c_rc_r	96·5±·54 (46)	95·9±0·68 (29)	94·2±0·45 (46)	94·8±0·55 (53)	90 8±0·49 (24)	91·7±0·54 (35)	87·8±0·80 (28)
zz	45·1±·21 (61)	43·9±0 24 (40)	43·9±0·21 (48)	47·4±0·21 (67)	43·4±0·27 (31)	42·9±0·25 (49)	45·8±0 32 (35)
c_yl	20·5±·19 (35)	19·8±0·20 (34)	20·4±0·18 (49)	21 4±0·18 (60)	19·0±0 18 (24)	18 0±0·16 (45)	19·4±0·28 (31)
ml	102·5±·56 (42)	104·1±0·64 (34)	102·5±0·62 (47)	108·1±0·34 (61)	98 5±0·57 (28)	99·4±0·61 (43)	102·4±0 61 (29)
c_pl	74 0±·34 (63)	74·9±0·40 (40)	74·4±0·45 (47)	83·2±0·44 (59)	67·7±0·37 (32)	69·7±0·37 (50)	75·9±0·55 (31)
rb'	32·2±·20 (63)	30 9±0·28 (40)	30·8±0·25 (49)	34·3±0·25 (65)	28·8±0 24 (32)	28·3±0·25 (50)	31·6±0·32 (34)
m_2p_1	28·0±·19 (47)	28 2±0·18 (22)	28·6±0·19 (34)	30 4±0·13 (63)	26·3±0·16 (19)	27·7±0·22 (19)	29·2±0·17 (31)
h_1	32·3±·25 (45)	30·9±0·46 (12)	33·4±0·34 (18)	33·3±0·29 (49)	28·9±0·45 (21)	29·7±0·36 (26)	30·8±0·32 (24)
m_2h	26·4±·16 (49)	24 9±0·45 (19)	25·8±0·32 (33)	26·3±0·21 (61)	23·4±0·34 (21)	23·6±0 51 (12)	24·5±0·31 (28)
c_rh	64·8±·43 (63)	64·8±0·47 (40)	65·8±0·51 (47)	64·0±0·45 (60)	57·2±0·46 (32)	56·5±0·47 (50)	55·9±0·56 (35)
rl	62·4±·47 (53)	62·2±0·40 (36)	62·5±0·42 (47)	62·9±0·45 (59)	57·0±0 42 (28)	53·5±0·44 (43)	56·3±0·65 (31)
$M\angle$	121°·5±·60 (63)	121°·7±0·60 (40)	120° 1±0·56 (47)	117°·0±0·58 (59)	126°·5±0·70 (32)	127°·8±0·59 (50)	121°·9±0·61 (31)
$R\angle$	71°·6±·75 (48)	72°·0±0·94 (37)	73°·5±0·77 (48)	70°·9±0·63 (58)	65°·8±0·79 (27)	71°·2±0·87 (48)	65°·2±0·71 (29)
$C'\angle$	70°·5±·66 (46)	61° 8±1·35 (16)	71°·2±1·20 (22)	78° 0±0·61 (40)	66°·4±0·99 (21)	67°·3±0·80 (27)	81°·6±0·78 (20)
$100\ c_rh/ml$	63·3±·50 (42)	62 4±0·66 (34)	64·4±0·61 (47)	59·3±0·49 (57)	58 0±0·50 (28)	56·8±0·49 (43)	54·4±0·59 (29)
$100\ c_rc_r/ml$	94 8±·71 (35)	92·3±1·03 (25)	92·0±0·61 (45)	87·8±0·63 (52)	92·1±0·70 (23)	92·4±0·74 (30)	85 8±0·94 (26)
$100\ g_og_o/c_pl$	130·6±·98 (63)	130·9±1·23 (40)	125·2±1·02 (47)	114·9±1·08 (55)	129·9±1·31 (32)	123 3±1·02 (50)	115·4±1·36 (23)
$100\ rb'/rl$	51·8±·48 (53)	50·0±0·54 (36)	49·3±0·69 (47)	54 7±0 44 (59)	50·8±0·52 (28)	53·2±0·66 (43)	56·5±0·75 (31)
$100\ g_og_o/c_rc_r$	100·7±·72 (46)	102·9±1·07 (29)	98·8±0·55 (45)	100·4±0·83 (48)	96·6±0·87 (24)	94·0±0·86 (35)	98·8±1·18 (22)

of the probability that the two samples compared were drawn from populations with identical means. The formula is theoretically more correct if the m characters used are uncorrelated with one another intra-racially. Ten were selected by Morant, principally because the correlations between them are low in the case of the male Qau series: of the total 45 r's 40 are less than ·3, the highest value is +·597 and no single character has more than two coefficients greater than ·3. The same 10 characters were used in calculating all the coefficients of racial likeness for mandibular measurements, including those given in the present paper. It may be noted that this number is considerably smaller than the 31 used, when possible, in the comparisons of cranial series by the same method.

Having the same nature as a measure of probability, the crude coefficient of racial likeness depends on the sizes of the samples compared. But the anthropologist is more interested in a measure of the absolute divergence of the types, and this is supposed to be obtained from the crude coefficients by adjusting them to values they might be expected to have if the samples were made up, not by the numbers actually available, but by 100 individuals each. If $\bar{n}_1$ and $\bar{n}_2$ are the mean numbers of individuals available for the m characters in the case of the first and second series in the comparison, respectively, then the reduced coefficient is defined to be:

$$50 \times \frac{\bar{n}_1 + \bar{n}_2}{\bar{n}_1 \bar{n}_2} \left\{ \frac{1}{m} \Sigma(\alpha) - 1 \right\} \pm 50 \times \frac{\bar{n}_1 + \bar{n}_2}{\bar{n}_1 \bar{n}_2} \times \cdot 67449 \sqrt{\frac{2}{m}}.$$

In a general way the classifications of groups of cranial series based on reduced coefficients of racial likeness that have hitherto been given accord with evidence of other kinds. The failure of the same method to give as reasonable results when applied to series of mandibles may possibly be due in this case to the inadequacy of one or other of the assumptions made in calculating the reduced coefficients. This possibility will be examined after presenting the results.

The reduced coefficients for the four new series, and between them and all the earlier ones, are given in Table X, and the values for all other pairs of the 17 male and 9 female series will be found in the papers by Morant and Martin. It has been shown repeatedly for cranial data that in attempting to derive a classification of the racial types from such material the most reasonable and suggestive results are always obtained if only the lowest orders of reduced coefficients are considered, while no account is taken of any greater than an arbitrarily defined limit. It appears to be an advantage to choose this limit as low as any which can be conveniently used for a particular group of series. In the case of the mandibular data the larger values of the reduced coefficients fail entirely to provide any arrangement of the types which could be supposed to indicate their inter-relationships and, accordingly, only the lower values will be considered now. In discussing the material available to him, Morant ignored all greater than 11, and for the material available now it was found that a limit of 10 could be used more

TABLE X

Reduced coefficients of racial likeness for mandibular series

Series	Sex	$\bar{n}$*	Spitalfields	Farringdon Street	Punjabi	Australian
Anglo-Saxon	♂	42·4	6·54±0·67	17·88±0·83	20·14±0·70	25·40±0·61
	♀	41·2	21·54±0·95	33·77±0·74	—	23·16±0·88
Dunstable	♂	37·3	6·57±0·72	11·82±0·88	15·81±0·75	26·26±0·66
Spitalfields	♂	48·5	—	2·82±0·79	5·46±0·66	38·10±0·56
	♀	26·0	—	16·75±0·95	—	51·07±1·09
Farringdon Street	♂	31·8	2·82±0·79	—	2·15±0·82†	49·56±0·73
	♀	40·3	16·75±0·95	—	—	41·99±0·89
Badari Egyptian (Predynastic)	♂	33·5	40·66±0·76	43·27±0·92	31·04±0·80	56·87±0·70
	♀	18·9	53·29±1·38	31·84±1·17	—	42·39±1·31
Qau Egyptian (4th–11th Dynasty)	♂	66·4	11·81±0·54	11·09±0·70	4·50±0·57	32·43±0·48
	♀	55·7	31·69±0·85	17·18±0·64	—	19·27±0·79
Sedment Egyptian (9th Dynasty)	♂	32·4	16·51±0·78	21·43±0·94	10·15±0·81	31·80±0·72
	♀	21·2	29·06±1·29	17·92±1·08	—	40·91±1·23
Kerma Egyptian (12th–13th Dynasty)	♂	55·7	24·65±0·58	29·34±0·74	18·28±0·62	19·75±0·52
	♀	44·7	50·39±0·92	29·36±0·71	—	29·78±0·85
Gizeh Egyptian *E* (26th–30th Dynasty)	♂	211·7	5·02±0·38	9·51±0·55	7·48±0·42	32·60±0·33
	♀	131·8	10·16±0·69	6·08±0·48	—	39·67±0·62
Tamil	♂	33·0	8·29±0·77	4·94±0·93	7·56±0·80	34·71±0·71
Punjabi	♂	43·4	5·46±0·66	2·15±0·82†	—	39·38±0·60
Nepalese	♂	18·9	13·26±1·11	10·67±1·27	10·91±1·15	26·77±1·05
Tibetan *A*	♂	24·9	12·41±0·92	8·72±1·08	17·01±0·96	30·25±0·86
Tibetan *B*	♂	11·9	26·17±1·58	36·08±1·74	32·49±1·61	9·71±1·52
Hylam Chinese	♂	38·8	5·24±0·70	14·33±0·86	17·89±0·74	26·39±0·64
Fukien Chinese	♂	37·5	18·22±0·71	30·09±0·88	27·15±0·75	22·16±0·66
Australian	♂	59·0	38·10±0·56	49·56±0·73	39·38±0·60	—
	♀	29·3	51·07±1·09	41·99±0·89	—	—

* The $\bar{n}$'s are the mean numbers of mandibles available for the 10 characters (w_1, zz, $c_y l$, ml, $m_2 p_1$, rb', h_1, rl, $R\angle$ and $100\, g_o g_o/c_y l$) used in computing the coefficients.

† The crude coefficient corresponding to this is 0·79 ± ·30.

conveniently. The arrangement suggested by the reduced coefficients can be appreciated most easily from Fig. 1, which shows all the connexions between the male series given by values less than 10. The Badari Predynastic Egyptian is the only series which has no reduced coefficient less than the arbitrary limit chosen.

The 17 series can be divided into three groups—an English, an ancient Egyptian and an Asiatic—and the Australian series. Considering these in turn, an unexpected relation is at once found in the insignificant coefficient between the Anglo-Saxon and Dunstable series, indicating that no distinction can be made

between their mandibular types although the cranial types are clearly differentiated (reduced coefficient = 13·66 ± ·45*). The mandibular value (− 0·18 ± ·76) is here less than the cranial and of an entirely different order. Equally unexpected is the linking of both the Anglo-Saxon and the Farringdon Street series to the Spitalfields, and also the lack of any connexion between the Anglo-Saxon and Farringdon Street series. The cranial evidence shows the inverse relation to this, viz. a close connexion between the Anglo-Saxons and seventeenth-century Londoners and a clear distinction between these two and the Spitalfields population, which is of uncertain date. Between the Anglo-Saxon and Farringdon Street series the cranial reduced coefficient is 8·79 ± ·32 and the mandibular 17·88 ± ·83; here the mandibular value is greater than the cranial and of a different order.

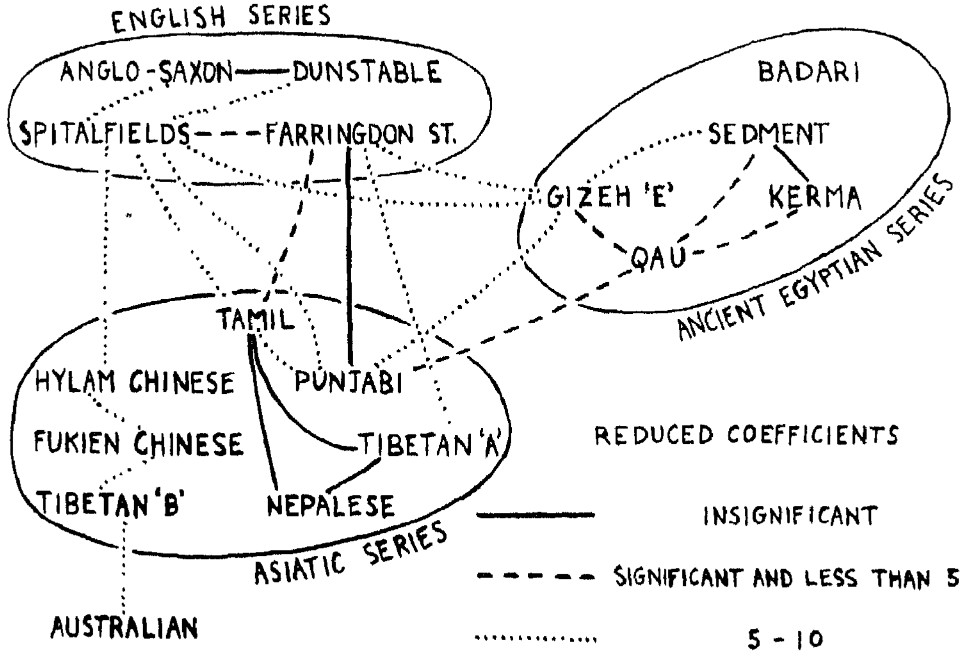


Fig. 1. The lowest reduced coefficients of racial likeness for 17 series of male mandibles.

Turning to the ancient Egyptian group, the fact that the Badari shows no connexion with any other series is not surprising, as it is the only one available of predynastic date and it is assigned to one of the earliest known predynastic periods. The insignificant coefficient between the Sedment and Kerma series (1·90 ± ·74) is unexpected, as the reduced coefficient for the cranial series is 16·41 ± ·31. For this Egyptian group, however, there are no results as unsatisfactory as those noted for the English group. The Asiatic mandibular series again show entirely unexpected resemblances and divergences. The Tamil, Nepalese

* The reduced coefficients of racial likeness for cranial series given here are all taken from papers in *Biometrika* published in 1926 or later.

and Tibetan A series have coefficients with one another which all differ insignificantly from zero, while the three corresponding cranial reduced values range from 12·6 to 40·9. For the Hylam and Fukien Chinese the reduced mandibular coefficient is $7{\cdot}66 \pm {\cdot}79$ which indicates distinct differentiation. The cranial types of these two series are very different, but it is suspected that this is due to the fact that the Hylam skulls were artificially deformed, though their facial and palatal measurements, which are not distinguished from the Fukien, do not appear to have been affected. The mandibles do distinguish the types, and this cannot be attributed to deformation.

These intra-group connexions do not encourage the hope that it will be possible to obtain any suggestive classification of the types from the reduced coefficients obtained from the mandibular measurements, and the comparison of series belonging to different groups makes this obvious. The most surprising connexion of the latter kind is the insignificant coefficient for the Farringdon Street and Punjabi series. But the Farringdon Street also has a lower value with the Tamil than with either the Anglo-Saxon or Dunstable series, and the Punjabi has a lower value with the Qau Egyptian than with any Asiatic series. It appears to be quite impossible to accept the coefficients as measures of racial relationship: they sometimes show close resemblance in type where no close racial affinity can be imagined, and they sometimes indicate clear distinction in type where close racial affinity must have existed. It may be noted that all the coefficients which make it impossible to obtain as suggestive a classification of the data as that obtained from the series previously dealt with are with one or other of three of the four new series. If the Spitalfields, Farringdon Street and Punjabi are omitted from Fig. 1 no connexions of the order considered are found between the three groups of series. The new material has apparently demonstrated the defect of the method.

It is clear that reduced coefficients of racial likeness for the mandible tend, in general, to be markedly lower than the values corresponding to them for the cranium, and only one case for which the reverse is true has been noted above, viz. that of the Anglo-Saxon and Farringdon Street series. This suggests that the unsatisfactory nature of the results shown in Fig. 1 may be due to the fact that some of the series used are too small for the purpose. The limiting size of cranial samples which yield suggestive and consistent results when compared in the same way has been determined empirically. In this case 50 is a safe limit to take, but samples composed of 30–50 individuals generally yield reliable results. The sizes of the series of mandibles can be judged from the $\bar{n}$'s in Table X, these being the average numbers of bones on which the means of the 10 characters used in computing the coefficients are based. Two of the male series have $\bar{n}$'s under 20, and no reliance whatever can be placed on results obtained from cranial samples no larger than these. If all series with $\bar{n}$'s less than 50 are ignored, we are only left with the Qau, Kerma and Gizeh (E) Egyptian and the Australian. The first

three are connected with one another and the last is widely removed from all of them by the coefficients, so there is nothing unexpected in these results, as far as they go. If the limit is lowered to 40, the Anglo-Saxon, Spitalfields and Punjabi series will also be included. Unexpected connexions are then found between the Punjabi series, on the one hand, and the Spitalfields, Gizeh Egyptian and Qau Egyptian, on the other, but of these three coefficients only one is less than 5 (Punjabi and Qau reduced $= 4{\cdot}50 \pm {\cdot}57$), and it has been shown that some ancient Egyptian and modern Indian cranial types are remarkably similar.* The results, which it is impossible to accept if the coefficients are considered as measures of racial relationship, are only evident when the shorter series are brought into the picture. The six insignificant coefficients, for example, are quite unexpected and unacceptable, but for every one of these one or both of the series compared has an $\bar{n}$ less than 40. The fact that several marked differences may be found between corresponding male and female coefficients in Table X, also suggests that several of the series are too short to give consistent results. The limiting size of sample required can only be determined empirically at present, as any theoretical estimate of it would require a knowledge of inter-racial variabilities which could only be found from far more extensive material than that available. It is quite possible that suggestive results would be given by series of mandibles made up by 50 or more individuals, or it may be necessary to adopt a still higher limiting size; and it may also be necessary to reject all reduced coefficients greater than 5, say, in interpreting such data. We cannot say that the method applied to measurements of series of mandibles is incapable of yielding results of value to the anthropologist, since it may be that the lack of suggestiveness of the arrangement shown in Fig. 1 is merely due to the fact that certain essential conditions were not observed in preparing that diagram. Data for additional series of a sufficient length will be required, either to justify the use of the coefficient of racial likeness in this case, or to demonstrate that it cannot be used profitably. All we can assert is that short series—composed of fewer than 40 mandibles, say—will not provide what is wanted.

Certain devices are used in calculating the reduced coefficients of racial likeness, and it may be suggested that these are partly responsible for the discordant results, and that the use of a theoretically more correct formula would modify them appreciably. The effect of the use of a single set of σ's instead of the values for each series used may be examined first. Martin has given reduced coefficients between the Egyptian E and a number of other series computed by using the Qau σ's, in one case, and those of the Egyptian E series itself, in the other.† Corresponding pairs of the lower coefficients are all in close agreement, though a few significant differences were found for the higher values which are

* See "A Study of the Badarian Crania recently excavated by the British School of Archaeology in Egypt", by Brenda N. Stoessiger, *Biometrika*, XIX (1927), pp. 110–50.

† *Loc. cit.* Table V.

neglected, however, in attempts to interpret the data. Three of the new male coefficients were calculated in the two ways with the following results, the first value of the reduced constant being that found by using the Qau σ's and the second that found by using the Egyptian E σ's: Farringdon Street and Punjabi $2{\cdot}15 \pm {\cdot}82$, $2{\cdot}56 \pm {\cdot}82$; Farringdon Street and Tamil $4{\cdot}94 \pm {\cdot}93$, $5{\cdot}21 \pm {\cdot}93$; Farringdon Street and Anglo-Saxon $17{\cdot}88 \pm {\cdot}83$, $22{\cdot}16 \pm {\cdot}83$. These results accord with Martin's, the coefficients calculated in the two ways being in close agreement in the case of the two lower pairs. It is unlikely that any of the unexpected relationships found between the series can be attributed to the use of a constant set of σ's in place of the sets for each of the series in a particular comparison.

It is unlikely, too, that the results obtained from the reduced coefficients of racial likeness differ appreciably from those which would be given by a theoretically more correct formula which takes into account the intra-racial correlations between the different measurements used. The 10 characters were chosen because the correlations between them are nearly all of a low order, and there is a far closer approach to the ideal condition here than in the case of the characters used in computing the cranial coefficients. Also, the mandibular coefficients which differ insignificantly from zero show a difference of means for nearly every character considered separately which would usually be considered insignificant,* and under these circumstances it cannot matter much whether the correlations between the characters are taken into account or not. These insignificant coefficients are largely responsible for our inability to accept the criterion as a measure of racial relationship.

Consideration of the same group makes it evident that the method of "reducing the crude coefficient cannot be responsible for all the unexpected results obtained. In the case of the six crude coefficients which differ insignificantly from zero there is, in fact, no need to reduce them, and it is clear (*from cranial evidence*) that the device used achieves the end in view sufficiently well in other cases. As far as can be seen now, therefore, no one of the assumptions made, or devices used, in applying the method of the coefficient of racial likeness can be considered responsible for the failure of the method to give results of value. This failure may be due to the fact that it has been applied to samples which are too small. Another possibility is that the group of measurements used is unsuitable for the purpose in view, and this is discussed in the following section.

8. *A comparison of single characters.* The relative values of different characters for purposes of racial classification can be estimated from the α's found in computing the coefficients of racial likeness. An α is approximately the square of a quantity which is the difference of two means divided by its standard error, and the difference—if considered by itself—may be supposed clearly significant if the α is greater than 10. Comparisons have been made between 17 male series for the same 10 characters, so there is a total of 1360 α's for these. Of this total 374

* See p. 109 below.

(27·5 per cent.) are greater than 10, and for 12 of the 17 series Morant found a percentage of 27·3.* These percentages are in remarkably close agreement, in spite of the fact that they depend to a certain extent on the sizes of the series compared: on the average, comparisons of longer series will be expected to show more significant α's than comparisons of shorter series. But it is clear that for the series available the characters used are capable of making clear distinctions. For each of the 10 characters, in the comparison of the 17 series, there is a total of 136 α's. The percentages of α's greater than 10 are: $R\angle$ 12·5, rl 20·6, h_1 24·3, c_yl 25·0, m_2p_1 25·7, zz 28·7, ml 28·7, rb' 30·9, 100 g_og_o/c_pl 36·8, w_1 41·9. Some characters evidently distinguish the types far more effectively than others, and it must be remembered that two were omitted from the list used in computing the coefficients because they appeared to be practically constant for the series considered by Morant. For the mandibular angle ($M\angle$) he only found three significant differences among 66 comparisons of mean values. This was the more surprising since anthropologists have often supposed that this character is of peculiar importance. It shows great intra-racial variability—the standard deviations for it being of the order 6°—and the means for 16 male series all lie between 120°·0 and 125°·3. It is true that the Australian mean of 117°·0 for 59 male mandibles is clearly divergent. The index expressing the breadth at the angles as a percentage of the breadth at the tips of the coronoid processes (100 g_og_o/c_rc_r) was also omitted because it only showed two significant differences in 66 comparisons. The intra-racial standard deviations for this character are of the order 7·0 and the range of the means for 17 male series is 95·0–102·9: the value of 100·4 for the Australian bones is not peculiar.

At the other extreme we find the mental angle ($C'\angle$) for which inter-racial variability is evidently much greater in proportion to intra-racial variability than in the case of the two preceding characters. This showed 35 significant differences among 66 comparisons of means, but it was not included among the characters used in computing coefficients of racial likeness because it was feared that it is a less reliable measurement than most of the others. It has been shown above (p. 89) that the readings of two observers sometimes show a very satisfactory agreement, and it is unlikely that personal equation is a disturbing factor in the case of comparisons between most of the means available for this character. A greater angle denotes a lesser projection of the chin. For 4 English male series the means range from 61°·8–70°·5, for 5 Egyptian from 70°·2–75°·5, for 7 Asiatic from 62°·9–77°·1 and the Australian mean for 40 bones is 78°·0. Several of these means are based on small numbers of specimens, and some have standard errors of the order 2°, but it is clear that the character often makes very clear distinctions between racial types. The Australian mean is extreme, but less removed from some of the others than would have been anticipated.

We may now ask whether the failure of the method of the coefficient of racial

* *Loc. cit.* p. 114.

likeness to give suggestive results when applied to series of mandibles is due to the choice of characters used in computing it. One of the most disconcerting results is the occurrence of insignificant coefficients in cases where distinct differentiation would have been expected, and where it is shown by the coefficients for the corresponding cranial series. Six insignificant values for male series of mandibles have been found (see Fig. 1), involving nine series. Standard deviations have only been given for four of these—the Dunstable and Farringdon Street English, the Punjabi and the Kerma Egyptian—as the others were considered too short for the purpose, and for the remaining five the Qau Egyptian constants may be applied, as in computing the coefficients. Comparisons of the means are summarized in Table XI, "none" signifying that there are no differences greater than three times their probable errors and the numbers in brackets being the ratios of the differences to their probable errors in cases where these are greater than 3.

TABLE XI

*A comparison of the significant differences between means for two groups of characters, in cases where the coefficients of racial likeness indicate an insignificant difference**

	10 C.R.L. characters	11 other characters
Anglo-Saxon and Dunstable	None	None
Farringdon Street and Punjabi	h_1 (4·4), 100 $g_o g_o/c_p l$ (3·6)	$g_o g_o$ (5·7), $C' \angle$ (5·2)
Sedment and Kerma Egyptian	ml (4·7)	100 $c_r h/ml$ (3·4), 100 $c_r c_r/ml$ (3 2), $C' \angle$ (3·9)
Tamil and Nepalese	h_1 (3·1)	$C' \angle$ (4·6)
Tamil and Tibetan *A*	None	$C' \angle$ (5·5)
Nepalese and Tibetan *A*	None	$C' \angle$ (3·3)

* See text for explanation.

In these 6 cases the 11 characters which are not used in computing the coefficients do tend to show a larger number of significant differences, and clearer differentiation in the case of some characters, than do the 10 characters used. But this difference depends almost entirely on the mental angle ($C' \angle$), and if it were omitted the choice of any group of characters from the remaining 20 would lead to almost identically the same results as those derived from the group adopted: there would be no clear distinction between the pairs of series compared. The situation is changed if $C' \angle$ is included, but it is unsuitable as a coefficient of racial likeness character, since it is feared that its readings for some of the earlier series were not found in precisely the same way as that employed later.

The Australian series shows no low coefficient with any other, and this is largely due to the fact that two of its means (for zz and $m_2 p_1$) are the greatest yet found. But the same series also has the greatest $c_p l$ and $C' \angle$ and its $M \angle$ is the

smallest, and these three characters are not used in computing the coefficients. The type would almost certainly be distinguished equally clearly if the 11 remaining characters were used for this purpose instead of the 10 chosen. There is no doubt that the coefficients would be changed to some extent if they were computed for a different set of characters, but it seems probable that their orders would be little affected, and that the unexpected results which make it necessary to question the utility of the method would still be found.

A comparison of characters considered singly throws some light on the cause of these unexpected results. It will be sufficient to consider the 4 English series, which give mandibular coefficients markedly different from those found for the corresponding, but longer, series of crania. There are 6 comparisons, based on 10 characters, and there are only 20 of the 60 α's greater than 4. An α is approximately the square of a quantity which is the difference of two means divided by the standard error of the difference, and it would be expected to show some values greater than 4 in a set of 60 comparisons merely as the result of chance, if in fact all the series represented the same population. Only 7 of the α's are greater than 8, the largest being 22·9 and the next largest 17·9. For these four series there are very few differences which are markedly significant. There is sufficient evidence to show that some pairs of the types do differ significantly, but it is also clear that the estimates of divergence in type provided by the coefficients are likely to be particularly unreliable owing to the influence of errors of random sampling. For the material available errors of this kind may be large enough to obscure the situation. This seems to be a possibility, and in view of it our general conclusion must be not that coefficients of racial likeness based on measurements of series of mandibles are incapable of revealing racial relationships, but that longer series than some of those used above will be needed in order to examine the use of the method applied to such material. Short series—made up by fewer than 40 bones, say—will certainly not give what is needed.

9. *Conclusions.* This paper presents the results of a statistical treatment of two English (male and female), a Punjabi (male only) and an Australian (male and female) series of mandibles. Measurements were taken in accordance with the biometric technique, and estimates of their accuracy were obtained by repeating a number and comparing the distributions of first and second readings. The two English series are not associated with individual crania or other parts of the skeleton, and the problem of sexing these is discussed. It is shown that a crude mathematical method and anatomical appreciation agree in about 85 per cent. of cases, and there is reason to believe that the sexes finally adopted give the same order of accuracy as those obtained by sexing a series of crania anatomically. The constants of variation reveal a few significant, but very small absolute, differences between the variabilities of the different series available, and this conclusion is the same as that derived from cranial measurements. At the same time the mandible tends to be rather more variable, relative to size, than the

cranium. Special topics discussed are intra-racial correlations of the measurements, asymmetry and records relating to the teeth lost before death.

Racial comparisons are made by using the method of the coefficient of racial likeness. In all there are 17 male and 9 female series which can be used for this purpose, though several of these are evidently too small to be of any permanent value by themselves. The coefficients show a number of entirely unexpected resemblances and divergences, and it is clear that they do not provide a rational classification of the types. In general they differentiate the series far less effectively than do the corresponding cranial coefficients. In 6 cases out of 136 comparisons there is no evidence of a significant difference judging from the mandibular measurements, although the series would be expected to represent quite distinct races and the corresponding cranial coefficients indicate clear divergence. It is shown that this result is not due to an unsuitable choice of the characters used, and that it cannot be attributed, as far as can be seen, to the disturbing influence of any of the assumptions made, or devices used, in computing the coefficients. The failure of the method may well be due to the fact that short series of mandibles are not capable of providing a reliable classification of the races they represent. The longer series available do give suggestive results, but there are not enough of them to suggest that additional long series will probably do the same. We can assert that series made up by 40 or fewer individuals will not give the information required, and for such the lack of statistical distinction between two types cannot be supposed sufficient evidence of racial identity. Series made up by 40–50 individuals may be sufficiently long, but in further investigations on the same lines it would be safer to exclude all composed of fewer than 50 bones. It is quite likely that it will be possible to demonstrate the utility of the method when it is applied to sufficiently long series.

I wish to thank Dr Morant for the photographs reproduced, and Miss A. B. Clements for typing the manuscript of this paper.

DESCRIPTION OF PLATES

Plates I, II and III show standard aspects of typical male mandibles, the focal plane of the camera having been perpendicular or parallel to the standard horizontal plane of the bone. In these cases a lens with a long focal length was used, and the distance from lens to object was about 2½ metres. The small images obtained were enlarged in printing, and the prints are reproduced here approximately at 0·9 natural size. Distortion may be considered negligible. The photographs reproduced in Plates IV and V were taken with a lens having a shorter focal length and at a closer distance. The typical male mandibles were selected by considering the deviations of the measurements of shape (angles and indices) for each bone of a series from the means for the series in terms of the standard deviations for each of these measurements. Each of the three bones has every index and angle differing from the mean for the series to which it belongs by less than 1·2 times the standard deviation of the measurement. Also, their maximum breadths (bicondylar, w_1), lengths (total projective, ml) and heights (projective height of coronoid process, $c_r h$) fall within the same range, except that the coronoid height of the selected Farringdon Street mandible differs from the mean for the male series by an amount which is 1·7 times the standard deviation of the distribution. Bones which are more typical than the three shown could not be found in the short

series available, but it should be realized that a comparison of these may suggest that there are differences in metrical characters, or anatomical details, which would not be found if truly typical specimens—i.e. ones representing the averages in all respects—were available.

Plate I Typical Punjabi (above, No. 6·3616) and Australian (No. 20·3003) male mandibles: *norma verticalis*. These two show little difference in size: for the true racial types the Australian seen from this aspect would show a rather larger excess in size over the Punjabi. There are clear differences in massiveness, and in the ways in which the teeth are set in the bones.

Plate II. Typical English (A, Farringdon Street, No. 622), Punjabi (B, No. 6·3616) and Australian (C, No 20·3003) male mandibles: *norma lateralis*. The mandibular angles are seen to be very close and the means for the three series are closer still. For the true types the breadth of the ramus relative to its length would distinguish the Australian from the other two rather more clearly than is the case for the selected specimens. The lesser projection of the chin is the most striking characteristic of the Australian mandible, and this is also characteristic of the series.

Plate III. Typical English (A, Farringdon Street, No. 622), Punjabi (B, No. 6·3616) and Australian (C, No. 20·3003) male mandibles: *norma frontalis*. The differences in size are small, but the Australian is clearly the most massive bone and the setting of the teeth in it is characteristic.

Plate IV. Contrasted forms of Australian mandibles.

A. Male bones with extreme mental angles; 0·7 natural size. The mandible on the left (No. 20·59) has the lowest mental angle ($C' \angle = 65°{\cdot}5$) for the series, and the one on the right (No. 20·6213) the highest (92°·5). The mean angle is 78°·0 and the typical male (Plate II C) has a reading of 81°·5.

B. Female bones with extreme mental angles: 0·8 natural size. The mandible on the left (No. 20·6202) has the lowest mental angle (71°·0), and the one on the right (No. 20·8461) the highest (94° 0).

C. The dental arcades of two male Australian mandibles of contrasted forms: 0·9 natural size. The specimen on the left (No. 20·6211) has a parabolic arch and that on the right (No. 20·8521) differs from it in having the front teeth (incisors and canines) almost in a straight line. The difference is seen to be dependent more on the inclinations of the incisors than on the positions of their sockets.

Plate V. Pathological and anomalous Australian mandibles.

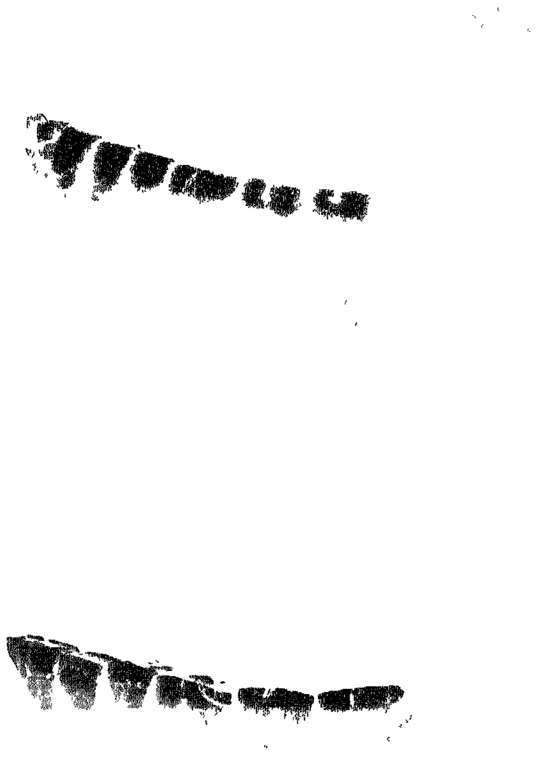
A. A female mandible showing marked erosion of the angles due to syphilis: No. 3955·2, 0·9 natural size.

B. A male mandible showing severe healed injury of the right ramus: No. 20·8562, 0·9 natural size.

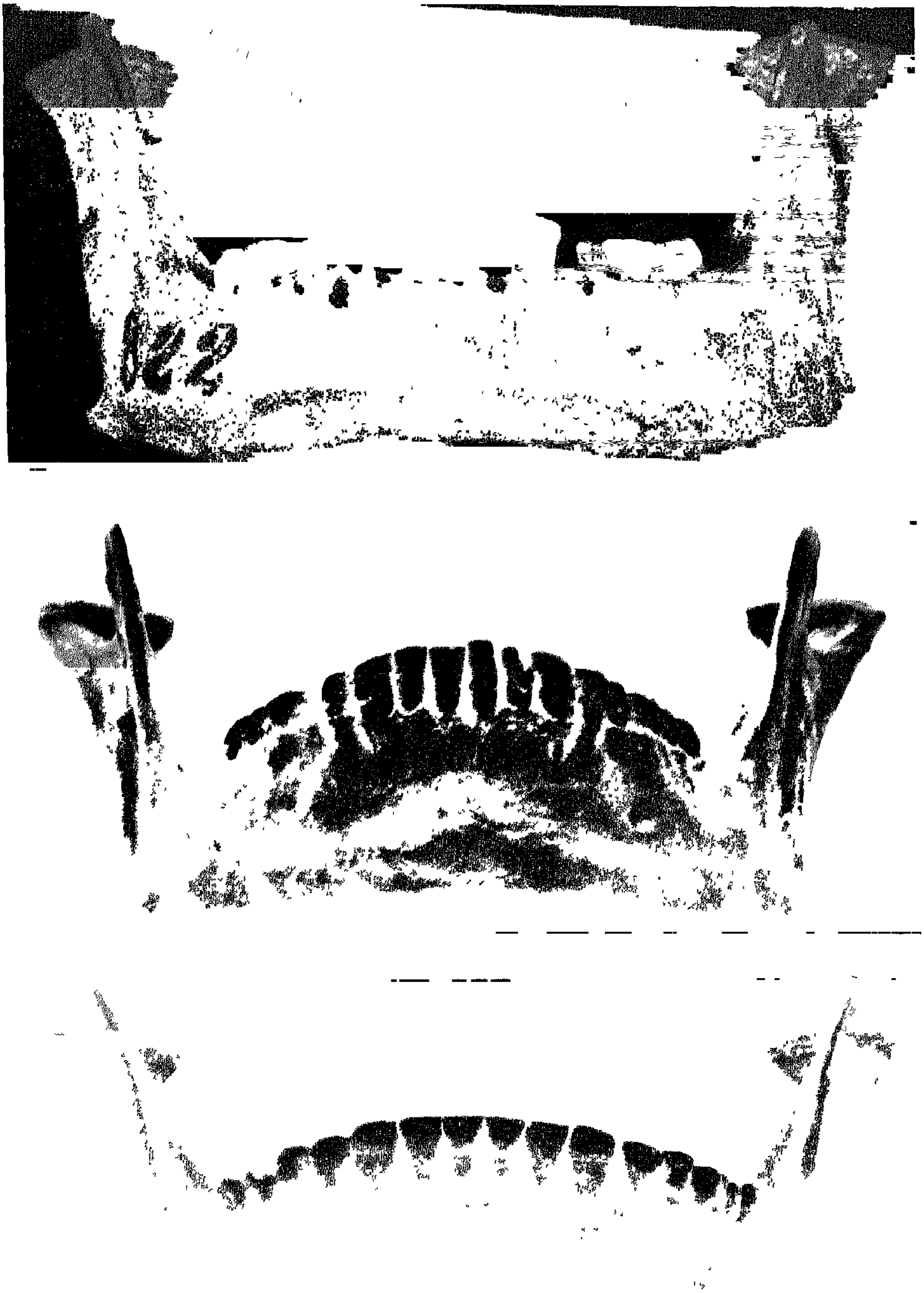
C A male mandible of a remarkably massive and primitive type: No. 20·8551, 0·9 natural size.

D. A male mandible showing gross overcrowding of the incisors: No. 20·7702, 1·3 natural size.

Typical Punjabi (above) and Australian Male Mandibles.



Typical English (A), Punjabi (B) and Australian (C) Male Mandibles.



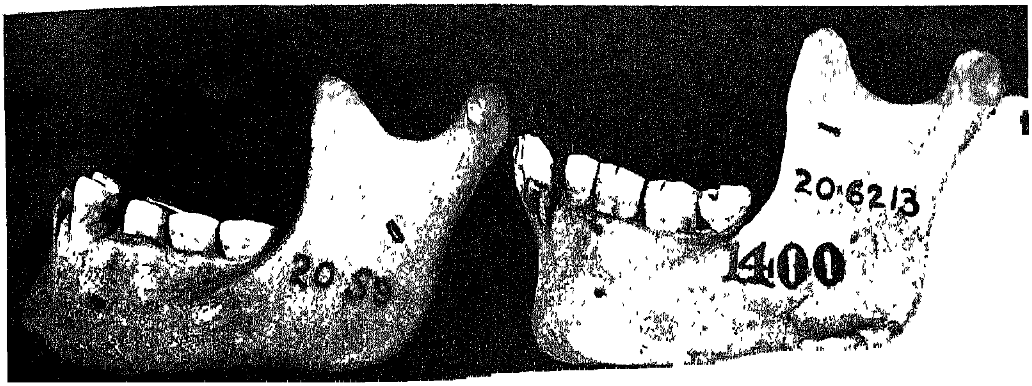
A

B

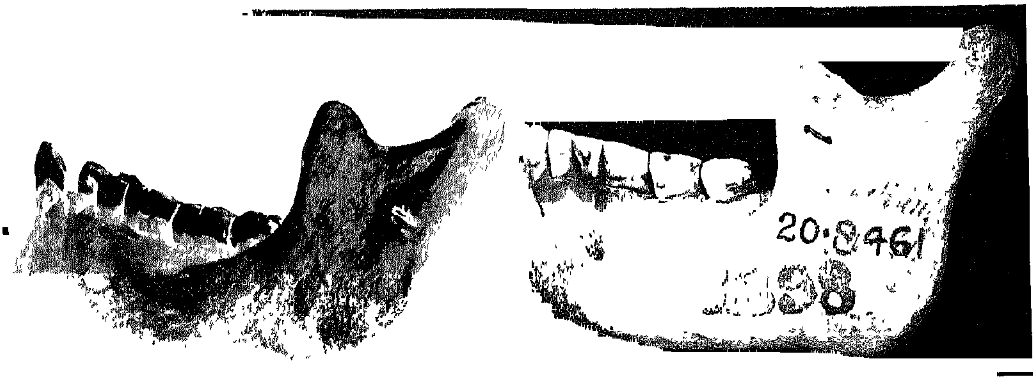
C

Typical English (A), Punjabi (B) and Australian (C) Male Mandibles.

Cleaver, *Biometric Study of the Human Mandible*



A. Male bones with extreme mental angles.

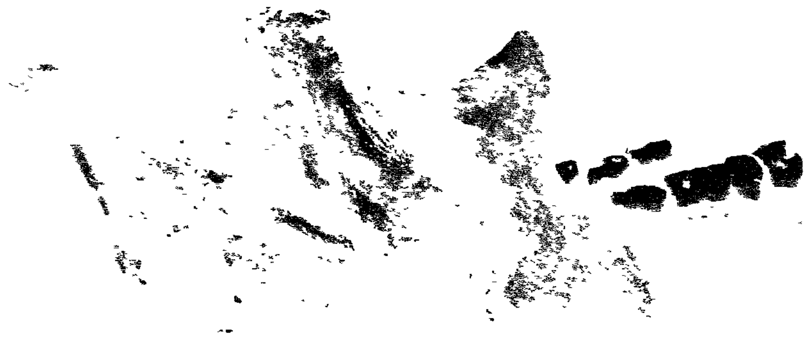


B. Female bones with extreme mental angles.

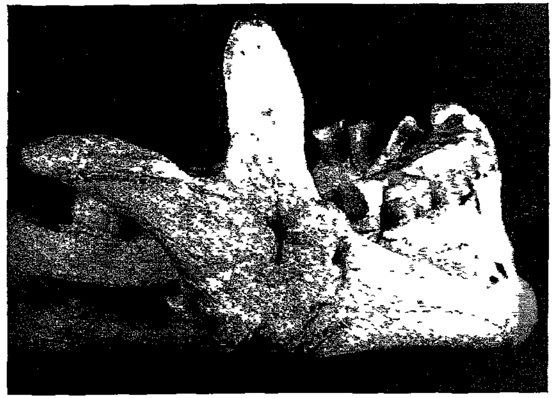
C. Different forms of the dental arcade

Contrasted Forms of Australian Mandibles.

Cleaver, *Biometric Study of the Human Mandible*



A Erosion of angles due to syphilis.



B Healed fracture of right ramus.

C A remarkably primitive type.

D. Gross overcrowding of incisors

A BIOMETRIC STUDY OF THE HUMAN MALAR BONE

BY T. L. WOO, PH.D.

1. *Introduction.* The measurements of the skull which have been most widely used for anthropological purposes were originally defined, or elaborated from definitions of earlier workers, by Paul Broca and a number of his German contemporaries. The French and German techniques were by no means identical, but the two sets of measurements corresponded in a general way. They aimed at giving a general description of the cranium considered as a whole and of all its principal parts. The framers of the techniques were primarily anatomists who had become interested in anthropological problems, but there are no peculiarly anatomical considerations underlying their systems. In particular there was, in the case of the majority of the measurements, an unfortunate disregard of the fact that the skull is made up of a considerable number of different bones. Nearly all the later craniometric techniques are based on the earlier ones, and their general aim has been to secure greater precision and standardization. The result has been that a particular set of measurements has been given in a large number of publications for some tens of thousands of skulls representing extinct and existing races in all parts of the world. The value of this corpus of material is beyond question and, in fact, it is by far the most valuable material available at present which can be used to estimate with precision the resemblances of different varieties of man. It is known that all the usual measurements make some clear distinctions when the averages for different series are compared: in other words, they are all of racial significance But it is also known that their relative values for the purpose of differentiating races differ greatly. Some appear to be almost constant for all races, while others usually show significant differences, and it is found that there is something like a gradual transition between these extremes. This grading of the characters, in order of their effectiveness as racial criteria, could not be appreciated until extensive data had been collected for them.

The position with regard to the customary measurements suggests that it should be possible to select a smaller number of characters which could be used as, or more, effectively for purposes of racial classification, with less labour involved in recording and computing. The list chosen might be made up partly by some of the old measurements and partly by new ones. It is generally recognized that certain features of the cranium which are obviously of value in aiding racial discrimination are not estimated by any of the classical measurements. New measurements of the "flatness" of the facial skeleton were taken on nearly 6000 skulls, representing a number of races from different parts of the

world, with the object of examining the value of one such feature.* It was concluded that a few of these are as useful for the purpose in view as any other characters that have been dealt with metrically, and far more useful than some for which extensive records are available.

The investigation described in the present paper was undertaken in the hope of discovering other new measurements which might be of exceptional value in aiding racial classification. In a paper published in 1931† the writer gave definitions of 25 chords and arcs, of which the majority were new, each being confined to a single bone of the skull. These were taken on a series—the Egyptian *E*—of nearly 900 male specimens obtained from a single cemetery at Gizeh which was used from the 26th to the 30th dynasties. Among the measurements dealt with in this study are two of the malar bone, the horizontal arc and the vertical arc. These were determined for both malar bones, so that the question of asymmetry could be investigated,‡ but no material was collected then to throw light on the possible sexual and racial significance of the measurements in question. One of the arcs was taken later by Dr von Bonin on a series of New Britain skulls,§ and the material for them presented below relates to an additional 710 crania, made up by 14 male and 2 female series representing races in different parts of the world. Two additional measurements of the malar bones of these 710 specimens were also recorded.

Anthropologists have hitherto devoted little attention to the malar bone, though a number of scattered remarks relating to racial differences in its size and form may be found in the literature. The earlier discussions of its metrical and anatomical variations are conveniently summarized by Le Double.‖ He remarks: "Il n'est pas démontré péremptoirement encore, par des mensurations multiples et précises, que le malaire ait, toutes choses égales d'ailleurs, des dimensions plus considérables dans une race que dans une autre et, dans une race quelconque, chez l'homme que chez la femme."

2. *The material measured.* All the skulls for which measurements of the malar bones are given for the first time in this paper are in the Museum of the Royal College of Surgeons, London. The writer measured them there in 1934 and he is greatly indebted to the authorities of the College, and particularly to Miss M. L. Tildesley, for granting him ready access to the specimens. The series are:

(i) *English.* 43 ♂. These came from a single cemetery at Portugal Street,

* T. L. Woo and G. M. Morant, "A Biometric Study of the 'Flatness' of the Facial Skeleton in Man", *Biometrika*, XXVI (1934), pp. 196–250.

† T. L. Woo, "On the Asymmetry of the Human Skull", *Ibid.* XXII, pp. 324–52.

‡ These malar bone measurements for the Egyptian series and the index derived from them are also treated by Karl Pearson and T. L. Woo in "Further Investigation of the Morphometric Characters of the Human Skull", *Ibid.* XXVII (1935), pp. 424–65.

§ "On the Craniology of Oceania. Crania from New Britain", *Ibid.* XXVIII (1936), pp. 123–48.

‖ *Traité des Variations des Os de la Face de l'Homme* (1906), pp. 114–65.

London and they are known as the King's College series. The skulls are probably of eighteenth-century date.

(ii) *French.* 28♂. Most of these came from the catacombs of Paris and they are all later than the Merovingian period. Measurements were only taken of the complete crania.

(iii) *Italian.* 92♂. These modern skulls came from 12 provinces in the northern and central parts of Italy.

(iv) *Egyptian: dynastic.* 26♂. These belong to middle and late dynastic times. Measurements were only taken of the complete crania.

(v) *Egyptian: Ptolemaic and Roman.* 31♂. Measurements were only taken of the complete crania.

(vi) *Negro: Nigeria.* 41♂. Of this total 34 skulls came from South Nigeria —representing mainly the Ibibio and Ekoi tribes of the Calabar region—and the others are from different parts of North Nigeria.

(vii) *Negro: Congo.* 36♂ and 21♀. The majority of these specimens represent the Batetela tribe who live near the Lubefu River.

(viii) *Hindu: Bihar and Orissa.* 36♂. Several castes of Hindus are represented, and the majority of the specimens came from the Patna district in the north-west of Bihar.

(ix) *Punjabi.* 80♂. These crania are of Mohammedans and several castes of Hindus.

(x) *Javanese.* 45♂. These came from various parts of Java and the neighbouring islands.

(xi) *Chinese.* 63♂. Nearly half of these specimens are known to have come from various localities on the south-east coast of China, and the majority of the others probably came from the south of the country.

(xii) *Eskimo.* 29♂. These came from various parts of Greenland and neighbouring islands to the west.

(xiii) *Maori: New Zealand.* 39♂. The majority of these specimens came from the North Island, principally from the vicinity of Auckland, but some are from unknown localities.

(xiv) *Kanaka:* 50♂ and 50♀. These specimens came from the Islands of Oahu and Hawaii, and the population of the former is better represented than that of the latter.

Every one of these 639 male and 71 female skulls is sufficiently complete to give all the measurements defined in the following section.

3. *Definitions of measurements of the malar bone.* Fig. 1 shows the left malar bone and surrounding regions of the facial skeleton: *FMT* is the point where the malar ridge crosses the fronto-malar suture, and this is practically the same as Martin's *fronto-malare temporale*; *ZM* is the lowest point on the malar-maxillary suture, so it is his *zygomaxillare*. The other two points used are

not defined by Martin. O is the point where the malar-maxillary suture crosses the lower margin of the orbit, and ZT is the lowest point on the zygomatic suture which is still on the lateral surface of the arch. The measurements are:

(*a*) Ml_1 = minimum horizontal arc from O to ZT.

(*b*) Ml_2 = minimum vertical arc from FMT to ZM.

(*c*) $100\ Ml_2/Ml_1$.

These three are the malar-bone measurements of the earlier studies of measurements of single bones of the cranium. They are available for the long Egyptian series of male skulls* and for all the new material. Ml_1 is also available for the new British series. The arcs, taken with a steel tape, are recorded to the nearest 0·5 of a mm.

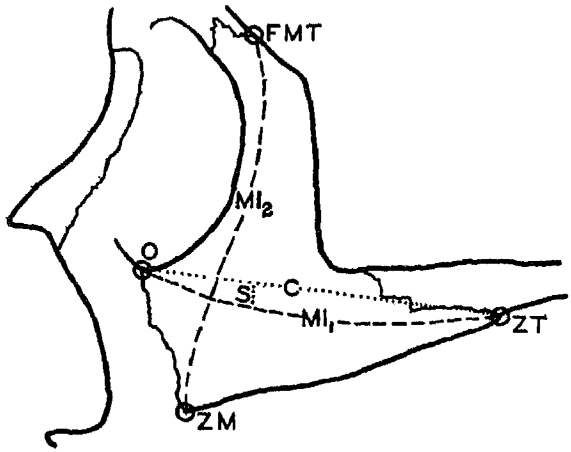


Fig. 1. The left malar bone and surrounding region, showing measurements taken.

(*d*) C = chord between the terminals of the horizontal arc (O and ZT).

(*e*) S = maximum subtense from this chord to the line marking the direction of the minimum horizontal arc (Ml_1). This line is first marked in pencil on the surface of the bone.

(*f*) $100\ S/C$. This provides a measure of the curvature of the horizontal arc.

These three measurements are only available for the new material.

The chord and the subtense were taken at the same time with the aid of a pair of co-ordinate callipers which was made for the writer by W. F. Stanley and Co. (London). This is similar in construction to the co-ordinate callipers made by P. Hermann, Rickenbach u. Sohn (Zürich), which could be used for the purpose, but the subtense arm of the new form terminates in a narrow straight

* An error was made in the tables of the asymmetry paper cited: the symbols Ml_1 and Ml_2 should be interchanged in these.

TA

Series
Eskimo
Chinese
Javanese
Kanaka
Maori
Negro: Nigeria
Negro· Congo
Egyptian: dynastic*
Egyptian: 26th–30th dynasties
Egyptian: Ptolemaic and Roman
Hindu: Bihar and Orissa
Punjabi
Italian
English
French
Kanaka
Negro: Congo

Series	Sex	No.
Eskimo	♂	29
Chinese	♂	63
Javanese	♂	45
Kanaka	♂	50
Maori	♂	39
Negro: Nigeria	♂	41
Negro: Congo	♂	36
Egyptian: dynastic*	♂	26
Egyptian: 26th–30th dynasties	♂	716, et
Egyptian: Ptolemaic and Roman	♂	31
Hindu: Bihar and Orissa	♂	36
Punjabi	♂	80
Italian	♂	92
English	♂	43
French	♂	28
Kanaka	♀	50
Negro: Congo	♀	21

Series
Eskimo
Chinese
Javanese
Kanaka
Maori
Negro: Nigeria
Negro: Congo
Egyptian: dynastic*
Egyptian Ptolemaic and Roman
Hindu: Bihar and Orissa
Punjabi
Italian
English
French
Kanaka
Negro: Congo

* New Em

Series	Sex	No.
Eskimo	♂	29
Chinese	♂	63
Javanese	♂	45
Kanaka	♂	50
Maori	♂	39
Negro: Nigeria	♂	41
Negro· Congo	♂	36
Egyptian: dynastic*	♂	26
Egyptian: Ptolemaic and Roman	♂	31
Hindu: Bihar and Orissa	♂	36
Punjabi	♂	80
Italian	♂	92
English	♂	43
French	♂	28
Kanaka	♀	50
Negro: Congo	♀	21

TABLE II. *Variabilities of malar-bone measurements*

Horizontal arc (Ml_1)				Vertical arc (Ml_2)				100 Ml_2/Ml_1	
σ		Coefficient of variation		σ		Coefficient of variation		σ	
R	*L*	*R*	*L*	*R*	*L*	*R*	*L*	*R*	*L*
6·00 ± ·53	5·21 ± ·46	8·51 ± ·76	7·33 ± ·65	3·81 ± ·34	3·73 ± ·33	7·25 ± 0·65	7·15 ± 0·64	5·54 ± ·49	5·80 ± ·51
4·07 ± ·25	3·86 ± ·23	6·49 ± ·39	6·14 ± ·37	2·66 ± ·16	2·82 ± ·17	5·36 ± 0·32	5·62 ± 0·34	6·29 ± ·38	5·95 ± ·36
4·18 ± ·30	4·00 ± ·28	6·82 ± ·49	6·58 ± ·47	3·48 ± ·25	3·66 ± ·26	6·91 ± 0·49	7·26 ± 0·52	5·80 ± ·41	6·83 ± ·49
3·52 ± ·24	3·74 ± ·25	5·65 ± ·38	5·98 ± ·40	2·65 ± ·18	2·66 ± ·18	5·32 ± 0·36	5·36 ± 0·36	4·96 ± ·33	5·18 ± ·35
4·29 ± ·33	3·83 ± ·29	6·86 ± ·53	6·13 ± ·47	3·17 ± ·24	3·61 ± ·28	6·36 ± 0·49	7·24 ± 0·56	5·31 ± ·41	5·60 ± ·43
4·40 ± ·33	4·65 ± ·35	7·24 ± ·54	7·55 ± ·57	2·76 ± ·21	3·10 ± ·23	5·61 ± 0·42	6·28 ± 0·47	5·58 ± ·42	6·22 ± ·46
4·89 ± ·39	4·55 ± ·36	8·35 ± ·67	7·70 ± ·62	2·77 ± ·22	3·18 ± ·25	5·80 ± 0·46	6·62 ± 0·53	6·13 ± ·49	6·35 ± ·50
4·53 ± ·42	3·71 ± ·35	7·76 ± ·73	6·32 ± ·59	3·10 ± ·29	3·11 ± ·29	6·56 ± 0·62	6·52 ± 0·61	5·18 ± ·48	5·25 ± ·49
4·38 ± ·08	4·55 ± ·08	7·37 ± ·13	7·64 ± ·14	3·11 ± ·05	3·21 ± ·05	6·30 ± 0·11	6·42 ± 0·11	5·85 ± ·10	6·02 ± ·11
3·99 ± ·34	3·90 ± ·33	6·83 ± ·59	6·61 ± ·57	3·01 ± ·26	3·28 ± ·28	6·32 ± 0·54	6·79 ± 0·58	5·86 ± ·50	6·40 ± ·55
2·99 ± ·24	3·09 ± ·25	5·38 ± ·43	5·47 ± ·44	2·87 ± ·23	2·97 ± ·24	6·26 ± 0·50	6·47 ± 0·52	5·53 ± ·44	5·23 ± ·42
4·32 ± ·23	4·37 ± ·23	7·51 ± ·40	7·61 ± ·40	2·70 ± ·14	3·34 ± ·18	5·70 ± 0·30	7·04 ± 0·38	6·56 ± ·35	6·55 ± ·35
3·70 ± ·18	3·83 ± ·19	6·28 ± ·31	6·49 ± ·32	2·95 ± ·15	3·07 ± ·15	6·15 ± 0·31	6·38 ± 0·32	5·40 ± ·27	5·64 ± ·28
3·91 ± ·29	3·76 ± ·27	6·65 ± ·49	6·36 ± ·46	2·40 ± ·18	2·96 ± ·22	4·82 ± 0·35	5·89 ± 0·43	4·18 ± ·30	5·02 ± ·36
3·55 ± ·32	3·99 ± ·36	6·14 ± ·56	6·90 ± ·62	3·01 ± ·27	3·40 ± ·31	6·23 ± 0·56	6·98 ± 0·63	6·39 ± ·58	6·28 ± ·57
3·75 ± ·25	3·94 ± ·27	6·47 ± ·44	6·73 ± ·46	2·62 ± ·18	2·69 ± ·18	5·68 ± 0·38	5·86 ± 0·40	3·92 ± ·26	4·56 ± ·31
4·02 ± ·42	4·10 ± ·43	7·03 ± ·74	7·16 ± ·75	2·77 ± ·29	2·40 ± ·25	5·96 ± 0·62	5·16 ± 0·54	5·62 ± ·59	5·40 ± ·56

Horizontal chord (*C*)	Subtense to chord (*S*)	100 *S*/*C*

edge instead of a sharp tip. This makes it possible to determine the maximum subtense by merely bringing the edge down on to the arc, marked in pencil on the bone, from which the subtense is taken: there is no need to repeat this operation several times in order to find the maximum reading, as is necessary when the form of callipers with a pointed subtense arm is used. Both scales of the new instrument have verniers attached, and the chords and subtenses were recorded to the nearest 0·1 of a mm.

4. *Sexual and bilateral comparisons.* The metrical material available for estimating sexual differences for the malar bone is very scanty. There is one series of 50 male and 50 female Kanaka skulls and another of 36 male and 21 female Congo-Negro. Means are given in Table I and variabilities in Table II. The numbers are too small to give reliable sex ratios (male mean/female mean) for the absolute measurements, but as far as can be seen these are not peculiar for cranial measurements. The eight constants range from 1·068 to 1·102 for the Kanaka series and from 1·010 to 1·032 for the Congo, but it would be unwise to assume from such slender evidence that the races are distinguished by their average sex differences. There is no suggestion that the true ratios are significantly different for different measurements, or for the right and left sides in the case of the same measurement. No clearly significant differences are found between the corresponding male and female indices, and it is clear that any sexual differentiation—apart from that in absolute size—which might be deduced from the measurements could only be revealed by data for larger samples.

The data which can be used to examine bilateral differences for male skulls are far more extensive. In estimating the significance of such constants the bilateral correlations have to be taken into account, and these are given in Table III for the long Egyptian and the two longest of the new series. In the case of the three comparisons which can be made, the Egyptian constant is greater than those for the other two series, and most of the amounts by which its value exceeds theirs are significant. For all six characters, however, no significant differences are found between the corresponding Italian and Punjabi correlations. It is commonly found for anthropometric material that the more homogeneous series give the higher correlations, and this relation is observed in the present case. There are no marked differences between the correlations for different characters, except that those for the subtense and the index involving the subtense tend to be lower than the others. It may be suggested that this is due to the fact that readings of the subtense—quite the smallest measurement—were not recorded in sufficiently small units to give reliable correlations. An examination of the constants for the Italian series shows that this is not the case, however. Its highest bilateral correlation is for the vertical arc. Readings of this measurement were taken to the nearest 0·5 of a mm. and the standard deviation for it is about 3 mm., so the unit of measurement is about one-sixth

of the σ. Readings of the subtense were taken to the nearest 0·1 of a mm. and the standard deviation for it is of the order 1·2 mm., so the unit of measurement is here about one-twelfth of the σ. Merely on account of the way in which the measurements were taken, it might hence be expected that the arc would show a lower bilateral correlation than the subtense, but actually the latter has the lower value.

Comparisons between the means and standard deviations for the right and left sides are summarized in Table IV, and the data there refer only to the fourteen new series. Treating each character separately, the numbers of series showing the right constant greater than the left, equality, and the left constant greater than the right are given, and also all the ratios of the differences to their probable errors which exceed 3·5. In calculating these ratios* the bilateral correlations used were: (i) in the case of Ml_1, Ml_2 and 100 Ml_2/Ml_1 those of the long Egyptian series in all comparisons except those for the Italian and Punjabi series, the appropriate correlations in Table III being used for these; (ii) in the case of C, S and 100 S/C, those of the Italian series in all comparisons except the Punjabi. It will be seen from Table IV that few markedly significant differences are found for any character: larger series than any dealt with there are generally needed to reveal beyond question the asymmetry in type of any cranial measurement. Considering all the series together, there is a clear suggestion that both the horizontal and vertical arcs of the malar bone tend to be larger in size on the left than on the right. This accords with the results obtained for the paired bones—approximately 800 in number—of the male Egyptian series E skulls for which both differences of means are significant and of the same sign ($L > R$). For the 50 male skulls from New Britain Dr von Bonin found the left mean of the horizontal arc 0·1 mm. greater than the right, though this difference is quite insignificant. The means of the other characters give no clear indication of asymmetry and it is curious that this should be so for the horizontal chord, since the arc which has the same terminals shows a different relation. The comparisons in Table IV for standard deviations suggest that variability on the left side exceeds that on the right in the case of Ml_2 and $100\,Ml_2/Ml_1$, but that there is no bilateral difference in variability in the case of the other characters. For the long Egyptian series the left standard deviation was found to be significantly in excess of the right in the case of both Ml_1 and Ml_2. Of the four absolute measurements, Ml_2 is the only one for which there is a suggestion of a bilateral difference in relative variability, measured by the coefficient of variation. For this character the left constant exceeds the right in the case of 12 of the 14 short series: the long Egyptian series shows differences of the same sign for both Ml_1 and Ml_2, but the former is significant and the latter is not.

* The formulae which have to be used are given on pp. 329 and 337 of the writer's paper on asymmetry cited above.

TABLE III. *Bilateral correlations of measurements of malar bones: male series*

	No.	Horizontal arc (Ml_1)	Vertical arc (Ml_2)	Horizontal chord (C)
Italian	92	·7327 ± ·0223	·9123 ± ·0118	·8708 ± ·0170
Punjabi	80	·8248 ± ·0241	·8498 ± ·0210	·9130 ± ·0126
Egyptian: 26th–30th dynasties	716, etc.	·9399 ± ·0029*	9219 ± 0035†	—

	No.	Subtense to chord (S)	100 Ml_2/Ml_1	100 S/C
Italian	92	6644 ± ·0393	·8017 ± ·0251	·5782 ± ·0468
Punjabi	80	·7541 ± ·0325	·7430 ± ·0338	·6970 ± ·0366
Egyptian: 26th–30th dynasties	716, etc.	—	·8806 ± ·0057‡	—

* For 718 skulls. † For 817 skulls. ‡ For 716 skulls.

TABLE IV. *Bilateral comparisons of constants for malar-bone measurements: fourteen male series*

	$R > L$		Equality	$L > R$	
	No. of cases	Significant differences	No. of cases	No. of cases	Significant differences
			Means		
Horizontal arc (Ml_1)	2	—	1	11	Negro. Nigeria (4·8), Egyptian: Ptolemaic and Roman (3·6), Hindu (7·5)
Vertical arc (Ml_2)	1	—	1	12	Chinese (4·3), Egyptian· Ptolemaic and Roman (3·9)
100 Ml_2/Ml_1	6	Eskimo (4 0)	0	8	—
Horizontal chord (C)	9	Maori (3 6)	1	4	—
Subtense to chord (S)	5	Negro: Nigeria (3 7)	2	7	—
100 S/C	6	—	2	6	—
			Standard deviations		
Horizontal arc (Ml_1)	8	Egyptian: dynastic (4·1)	0	6	—
Vertical arc (Ml_2)	1	—	0	13	Punjabi (5·1), English (4·8)
100 Ml_2/Ml_1	4	—	0	10	English (3·7)
Horizontal chord (C)	6	—	0	8	—
Subtense to chord (S)	8	—	0	6	Maori (4·2)
100 S/C	8	—	0	6	—

Bilateral comparisons of three constants have been made in Tables III and IV on the assumption that all races tend to show the same asymmetry in type or variability. The validity of this assumption may be examined. In the case of any particular one of the constants, it is clear that in the vast majority of cases the bilateral difference for a series A and the corresponding difference for another series B will be found to differ insignificantly from one another. By selecting extreme values, however, some significant differences of differences might be found. An examination of Table IV suggests that the clearest evidence of a racial difference in asymmetry, if such exist, is most likely to be shown in the case of the means for the horizontal arc (Ml_1). At one extreme the Hindu series has a mean for the left side which is 0·9 greater than that for the right, and the probable error of this difference is found to be 0·120 on the assumption that the bilateral correlation is the same as that for the long Egyptian series. At the other extreme the Javanese series has a mean for the *right* side which is 0·5 greater than that for the left and the probable error of this difference is found to be 0·144, on the same assumption. The difference of the differences for the two series is 1·4, and this is 7·5 times its probable error (0·187). This appears to afford clear evidence of racial distinction, but it is somewhat uncertain owing to the fact that the Egyptian bilateral correlation may differ appreciably from the unknown Hindu and Javanese values. And it must also be remembered that the case considered is an extreme value in a series of differences, so that a higher ratio of the constant to its probable error must be taken to indicate significance than would be the case if a single difference were being considered by itself. It will be safest to conclude that racial differences in asymmetry are certainly very small, and more abundant material would be needed in order to demonstrate beyond question that any races are differentiated in this way in the case of characters of the malar bone.

5. *The value of the measurements for the purpose of racial classification.* The measurements were taken with the primary object of discovering whether their averages for different series provide suggestive arrangements which might aid attempts to determine the racial relationships of the populations represented. The number of series measured is not large, but it should be sufficient to show which of the characters of the malar bone are likely to be most useful for the purpose in view. The means are given in Table I and the arrangements provided by three different pairs of the characters are shown in Figs. 2–4. In considering these figures it is necessary to appreciate in a general way the differences for each variate which may be taken to indicate clear differentiation.

Fig. 2 shows the inter-racial correlation of the horizontal and vertical arcs, the points being determined by the male means for the left side. In the case of both of these characters most of the differences between the means are large enough to indicate statistical significance. For the Punjabi and Hindu means in the left-hand bottom corner of the diagram, for example, the difference in the

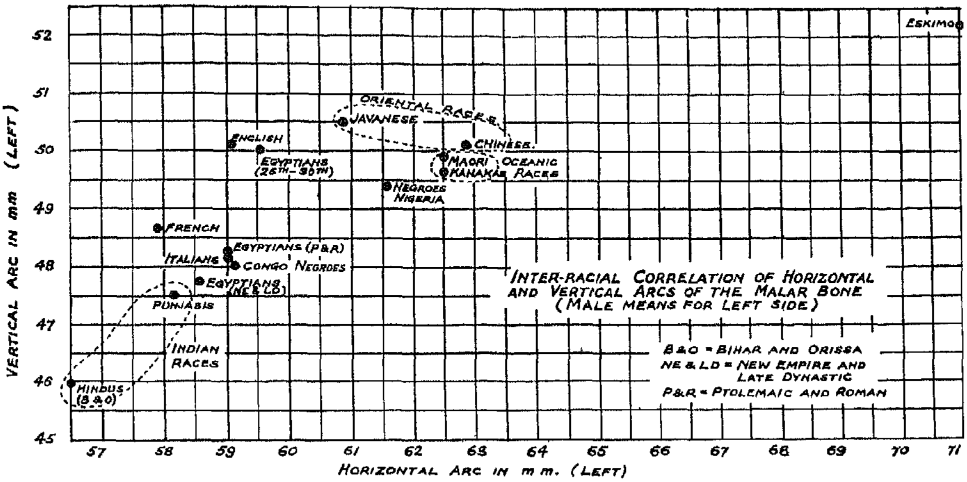


Fig. 2.

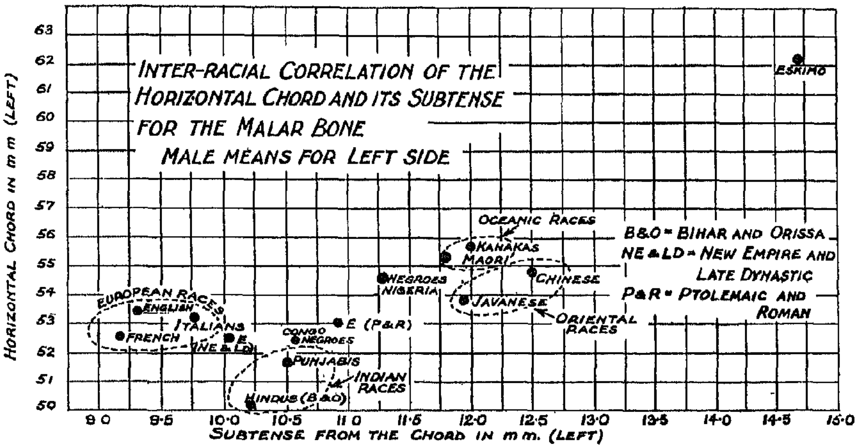


Fig. 3.

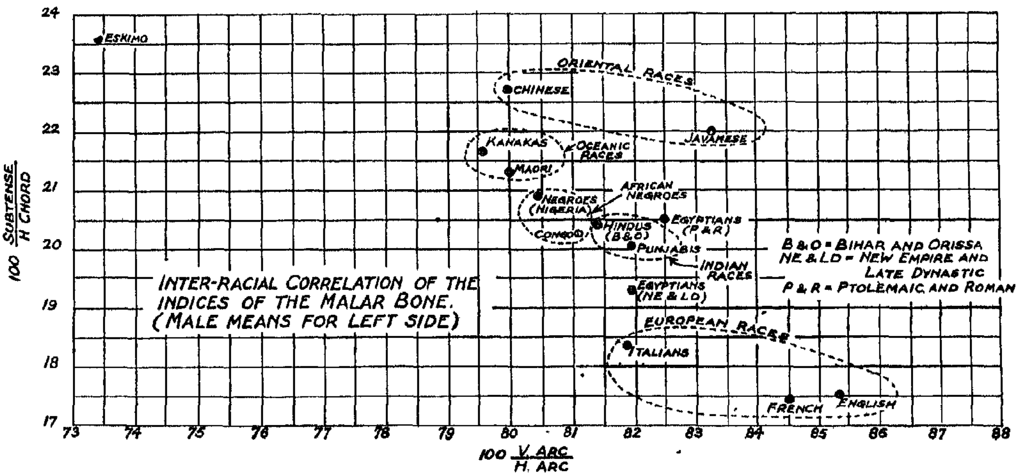


Fig. 4

case of the horizontal arc is 3·3 times its probable error, and that for the vertical arc 3·8 times its probable error. Both measurements are capable of making many clear distinctions between the racial types, and there is a suggestion that different members of the same family of races show small differences, while most of these families are distinguished from one another by occupying different areas of the bi-variate distribution. More abundant material would obviously be required to substantiate these points. The inter-racial distribution of either arc considered singly appears to be fairly continuous if the Eskimo series is omitted.* This has means widely removed and differing with marked significance from those for all the other series, and the large size of its malar bones appears to be a salient characteristic of the specialized Eskimo type. It should be noted that the measurements are not available for any American-Indian series. The two arcs appear to be highly correlated inter-racially, and the point for the Eskimo series is clearly the one which would be farthest removed from the regression straight line. Owing to the high correlation, we may anticipate that the index expressing the vertical as a percentage of the horizontal arc will be of little interest.

Fig. 3 shows the distribution of the series given by their means for the horizontal chord, and the maximum subtense to this chord. It is again found that most of the differences between the points are statistically significant in the case of each variate, and the Eskimo means again differ from all the others with marked significance. Some of the families of races appear to be separated, as before, and the differences between different races belonging to the same family are apparently all small. The inter-racial correlation between the chord and the subtense seems to be sensible but appreciably lower than that between the two arcs.

Fig. 4 shows the arrangement provided by the two indices. As was anticipated, the index $100\, Ml_2/Ml_1$ fails to arrange the series in any suggestive order. It is distinguished from all the absolute measurements by the fact that it shows far more insignificant than significant differences in the comparison of all possible pairs of the means, though the Eskimo mean still diverges widely from all others available. The index derived from the horizontal chord and the maximum subtense to it distinguishes the types far more clearly: in this case the Eskimo mean is still extreme, but it differs insignificantly from the Chinese which is nearest to it. This measurement of curvature appears to distinguish the different families of races rather more effectively than any one of the four absolute measurements does.

A provisional estimate of the value of the malar-bone measurements for anthropological purposes can now be given. The material available for them is ample enough to show that racial types of cranium have malar bones which differ very appreciably in both size and shape. Of the four absolute measure-

* The mean of 63·3 for the left horizontal arc given by Dr von Bonin for 50 New Britain skulls is close to the Maori and Kanaka means.

ments and two indices considered, the index obtained by expressing the vertical as a percentage of the horizontal arc is quite the most constant. This appears to be of little value for purposes of racial classification. The other five characters seem to differentiate the racial types as effectively as most of the usual cranial measurements, and more effectively than several of these. The material available suggests that they are characters which tend to be constant for races belonging to the same family of races, but which provide suggestive orders when the different families are compared with one another. They are thus of the same nature as skin colour, the nasal index, measures of prognathism and of the "flatness" of the facial skeleton, and indices derived from the lengths of the limb bones. Stature, the cephalic index and most calvarial measurements differ from these as they fail to make clear distinctions between the different families of races. Among races of the Old World the European and Indian, with their smaller and flatter malar bones, are at one extreme of the range: Oriental races have the largest and most curved bones, and negro and ancient Egyptian occupy intermediate positions. This arrangement accords in a general way with those provided by the indices measuring the "flatness" of the facial skeleton as a whole. In both cases, too, the Eskimo type has been found to occupy markedly aberrant positions. Judging from the short series measured, its malar bones are far larger than those of European, African, Asiatic and Oceanic types; they also show a greater degree of curvature, but the Eskimo type is most clearly distinguished by the fact that the heights of its malar bones (measured by the vertical arcs) are most peculiarly small compared with their antero-posterior lengths (measured by the horizontal arcs). This is of particular interest since the index measuring this ratio is the character least capable of distinguishing the other races from one another. The Eskimo type is detached, as it were, from the continuous system to which the others belong. It is generally recognized to be peculiarly specialized, but none of its characters are known to be more characteristic than these malar-bone measurements and the indices of facial "flatness". More of these data for Eskimo, Eastern Asiatic and American-Indian cranial series would probably throw as much light on the question of the affinities of the Eskimo population as any other new material.

While they are incapable of providing by themselves any reliable classification of the races of modern man, there is every promise that the malar-bone measurements dealt with in this paper will prove to be a valuable aid for the purpose when considered in conjunction with other characters. Hence it is suggested that they might be included with advantage in the routine descriptions of racial series of crania.

THE SAMPLING DISTRIBUTION OF THE CRITERION λ_{H_1} WHEN THE HYPOTHESIS TESTED IS NOT TRUE

[EDITORIAL NOTE. The criterion λ_{H_1} is appropriate to test the statistical hypothesis that the standard deviations of a character x are the same in a number, say k, of different normal populations. In the form $L_1 = \lambda_{H_1}^{2/N}$ (where N is the number of observations in the pooled samples), the criterion becomes the ratio of the weighted geometric to the weighted arithmetic mean of the k sample variances. For the special case where the samples are of equal size, tables of 5 % and 1 % probability levels have been determined by an approximate method by Mr P. P. N. Nayer.*

It is important however not only to have available these significance levels and so to control the risk of rejecting the hypothesis tested when it is true, but also to have some means of appreciating the chance that the test will detect real differences in population standard deviations when they exist. By this means it becomes possible to compare the efficiency of this and alternative tests. Over a year ago Dr S. S. Wilks promptly responded to a request of mine for help in this matter by providing the sampling moments of L_1^{-1}, in the general case where the population standard deviations are unequal. Since then Miss C. M. Thompson has compared his suggested Type III curves, having these moments, with a series of values of L_1^{-1} calculated from experimental sampling data. The correspondence between experiment and the Wilks's curves is excellent. Some further research into the matter is in progress. E.S.P.]

I. NOTE ON THE GENERAL SAMPLING MOMENTS OF λ_{H_1}

BY S. S. WILKS

Neyman & Pearson† have considered in some detail the problem of deriving a criterion λ_{H_1} for testing the hypothesis H_1 that k samples have come from populations with equal variances but with means having any values whatever. They have discussed the sampling theory of λ_{H_1} when H_1 is true. Here we shall be concerned with the more general case in which H_1 is not true. In order to

* *Statistical Research Memoirs* (Dept. of Statistics, University College, London), I (1936), p. 51. The substantial accuracy of the approximation involved has since been verified by Mr U. S. Nair whose work on the subject will be published shortly.

† J. Neyman & E. S. Pearson, "On the Problem of k Samples," *Bull. int. Acad. Cracovie*, Sér. A, 1931, pp. 460–81.

indicate more clearly the point of departure for this note we shall briefly describe what has been done.

Let the tth sample ($t=1, 2, \ldots, k$) of n_t individuals be denoted by Σ_t and suppose Σ_t has been drawn at random from a normal population with mean a_t and variance σ_t^2. Let $\bar{x}_t$ and s_t^2 be the mean and squared standard deviation of Σ_t.* In the present problem $\bar{x}_t$ and s_t^2 ($t=1, 2, \ldots, k$) are the only functions of the individual observations with which we shall be concerned. The probability that $\bar{x}_t$ and s_t^2 will fall in the infinitesimal ranges $\bar{x}_t \pm \frac{1}{2}d\bar{x}_t$, $s_t^2 \pm \frac{1}{2}ds_t^2$ ($t=1, 2, \ldots, k$) will be proportional to

$$C = \prod_{t=1}^{k} \frac{\left(\dfrac{n_t}{2\sigma_t^2}\right)^{\frac{n_t}{2}} (s_t^2)^{\frac{n_t-3}{2}}}{\sqrt{\pi}\,\Gamma\left(\dfrac{n_t-1}{2}\right)} e^{-\frac{n_t}{2\sigma_t^2}\{s_t^2+(\bar{x}_t-a_t)^2\}} \qquad \ldots\ldots(1)$$

We may regard the k samples $\Sigma_1, \Sigma_2, \ldots, \Sigma_k$ described by the k pairs of quantities $\bar{x}_t$, s_t^2 as having been drawn from the grand population (1).

Now H_1 is the hypothesis that the Σ_t are from populations with

$$\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2.$$

The set Ω of admissible populations consists of all populations (1) which could be obtained by taking all possible values of σ_t^2 and a_t. The set ω of populations is that subset of Ω for which the σ's are equal. The criterion λ_{H_1} is the ratio

$$\frac{C(\omega_{\max})}{C(\Omega_{\max})},$$

where $C(\omega_{\max})$ is the maximum of C taken over all populations in ω and $C(\Omega_{\max})$ is the maximum of C taken over all populations in Ω. Expressed in terms of the s's

$$\lambda_{H_1} = \prod_{t=1}^{k} \left(\frac{s_t^2}{s_a^2}\right)^{\frac{n_t}{2}}, \qquad \ldots\ldots(2)$$

where $$Ns_a^2 = \sum_{t=1}^{k} n_t s_t^2, \quad N = n_1 + n_2 + \ldots + n_k.$$

Neyman & Pearson† have considered the sampling properties of λ_{H_1} under the assumption that H_1 is true, that is, that the samples are drawn from a member of ω. We shall consider the sampling properties of λ_{H_1} under the assumption that the samples are from any member of Ω, in which the σ's are not necessarily the same.

Since we are using λ_{H_1} as an instrument for ordering the data embodied in $\Sigma_1, \Sigma_2, \ldots, \Sigma_k$ with respect to the tenability of H_1 it is clearly immaterial whether λ_{H_1} or any single-valued function of λ_{H_1} be used. From a theoretical point of

* Here s_t^2 is defined by $n_t s_t^2 = \sum_{i=1}^{n_t} (x_{ti} - \bar{x}_t)^2$.

† *Loc. cit.* pp. 467–73.

view the problem can be somewhat simplified by using $\lambda_{H_1}^{-2/N}$. Since λ_{H_1} is a function of $s_1^2, s_2^2, \ldots, s_k^2$, the gth moment of $\lambda_{H_1}^{-2/N}$ will be defined by the expression

$$\mu_g' = \int_0^\infty \cdots \int_0^\infty \lambda_{H_1}^{-2g/N}\, df_1 \ldots df_k, \qquad \ldots\ldots(3)$$

where

$$df_t = \frac{\left(\dfrac{n_t}{2\sigma_t^2}\right)^{\frac{n_t-1}{2}} (s_t^2)^{\frac{n_t-3}{2}}}{\Gamma\left(\dfrac{n_t-1}{2}\right)} e^{-\frac{n_t s_t^2}{2\sigma_t^2}}\, ds_t^2. \qquad \ldots\ldots(4)$$

Now it is clear from (2) that μ_g' will be the gth derivative with respect to θ of the following expression at $\theta = 0$,

$$\begin{aligned}\phi_g(\theta) &= \int_0^\infty \cdots \int_0^\infty (s_1^2)^{-\frac{n_1 g}{N}} \ldots (s_k^2)^{-\frac{n_k g}{N}} e^{s_a^2 \theta}\, df_1 \ldots df_k \\ &= K \prod_{t=1}^{k} \left[\frac{\Gamma\left(\dfrac{n_t-1}{2} - \dfrac{n_t g}{N}\right)}{\left(\dfrac{n_t}{2\sigma_t^2} - \dfrac{n_t}{N}\theta\right)^{\frac{n_t-1}{2} - \frac{n_t g}{N}}}\right], \qquad \ldots\ldots(5)\end{aligned}$$

where

$$K = \prod_{t=1}^{k} \frac{\left(\dfrac{n_t}{2\sigma_t^2}\right)^{\frac{n_t-1}{2}}}{\Gamma\left(\dfrac{n_t-1}{2}\right)}.$$

For $g = 1, 2$ we find with little difficulty

$$\mu_1' = \prod_{t=1}^{k} \left[\frac{\Gamma\left(\dfrac{n_t-1}{2} - \dfrac{n_t}{N}\right)}{\Gamma\left(\dfrac{n_t-1}{2}\right)} \left(\frac{n_t}{\sigma_t^2}\right)^{\frac{n_t}{N}}\right] \left[\sum_{t=1}^{k} \left(\frac{n_t-1}{2N} - \frac{n_t}{N^2}\right)\sigma_t^2\right], \qquad \ldots\ldots(6)$$

$$\begin{aligned}\mu_2' = \prod_{t=1}^{k} &\left[\frac{\Gamma\left(\dfrac{n_t-1}{2} - \dfrac{2n_t}{N}\right)}{\Gamma\left(\dfrac{n_t-1}{2}\right)} \left(\frac{n_t}{\sigma_t^2}\right)^{\frac{2n_t}{N}}\right] \left[\left\{\sum_{t=1}^{k} \left(\frac{n_t-1}{2} - \frac{2n_t}{N}\right)\frac{\sigma_t^2}{N}\right\}^2\right. \\ &\left. + \sum_{t=1}^{k} \left(\frac{n_t-1}{2} - \frac{2n_t}{N}\right)\frac{\sigma_t^4}{N^2}\right]. \qquad \ldots\ldots(7)\end{aligned}$$

In the important practical case in which $n_1 = n_2 = \ldots = n_k = n$, (6) and (7) reduce to

$$\mu_1' = \left[\frac{\Gamma\left(\dfrac{n-1}{2} - \dfrac{1}{k}\right)}{\Gamma\left(\dfrac{n-1}{2}\right)}\right]^k \left(\frac{n-1}{2k} - \frac{1}{k^2}\right)\left(\frac{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_k^2}{(\sigma_1^2\sigma_2^2 \ldots \sigma_k^2)^{\frac{1}{k}}}\right), \qquad \ldots\ldots(8)$$

$$\mu_2' = \left[\frac{\Gamma\left(\frac{n-1}{2}-\frac{2}{k}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\right]^k \left[\left(\frac{n-1}{2}-\frac{2}{k}\right)^2\left(\frac{\sigma_1^2+\sigma_2^2+\ldots+\sigma_k^2}{k(\sigma_1^2\sigma_2^2\ldots\sigma_k^2)^{\frac{1}{k}}}\right)^2 + \frac{\left(\frac{n-1}{2}-\frac{2}{k}\right)(\sigma_1^4+\sigma_2^4+\ldots+\sigma_k^4)}{k^2(\sigma_1^2\sigma_2^2\ldots\sigma_k^2)^{\frac{2}{k}}}\right]. \qquad \ldots\ldots(9)$$

When the hypothesis H_1 is true, that is, when $\sigma_1^2=\sigma_2^2=\ldots=\sigma_k^2$, it can be easily verified that

$$\mu_g' = \frac{\Gamma\left(\frac{N-k}{2}\right)}{\Gamma\left(\frac{N-k}{2}-g\right)} \prod_{t=1}^{k}\left[\left(\frac{n_t}{N}\right)^{\frac{n_t g}{N}} \frac{\Gamma\left(\frac{n_t-1}{2}-\frac{n_t g}{N}\right)}{\Gamma\left(\frac{n_t-1}{2}\right)}\right], \qquad \ldots\ldots(10)$$

as obtained by Neyman & Pearson. μ_g' will exist for all values of g for which the arguments of the gamma functions are positive.

The higher moments of $\lambda_{H_1}^{-2/N}$ become more and more complicated so that there is little hope of finding a workable form of the exact distribution function of $\lambda_{H_1}^{-2/N}$. Therefore, since $\lambda_{H_1}^{-2/N}$ has the range 1 to ∞, it appears that its distribution could be reasonably approximated by fitting a Type III curve by means of the first two moments. Let the form of the curve be

$$\frac{e^b b^a}{\Gamma(a)}(x-1)^{a-1}e^{-bx}. \qquad \ldots\ldots(11)$$

Equating the first two moments of (11) about the origin to μ_1' and μ_2' and solving for a and b, we find

$$\left.\begin{aligned} a &= \frac{(\mu_1'-1)^2}{\mu_2'-\mu_1'^2} \\ b &= \frac{\mu_1'-1}{\mu_2'-\mu_1'^2} \end{aligned}\right\}. \qquad \ldots\ldots(12)$$

II. AN INVESTIGATION INTO THE ADEQUACY OF DR WILKS'S CURVES

BY CATHERINE M. THOMPSON

The basic data used consisted of 500 samples of (i) $n_1 = 5$, (ii) $n_2 = 10$ and (iii) $n_3 = 15$ from a common normal population, obtained with the help of Tippett's Random Numbers. The values of, say, $v = \Sigma(x-\bar{x})^2/n$ had already been calculated for each of the 1500 samples for another purpose. By multiplying the values of v by appropriate factors it was possible to obtain 500 sets of values s_1^2, s_2^2 and s_3^2 from populations having unequal variances σ_1^2, σ_2^2 and σ_3^2, respectively.

For example, if v_1, v_2 and v_3 denote the basic sample variances, if the common population standard deviation is unity and if $N = n_1 + n_2 + n_3 = 30$, then

$$L_1 = \frac{\prod_{t=1}^{3} (s_t^2)^{\frac{n_t}{N}}}{\frac{1}{N} \sum_{t=1}^{3} (n_t s_t^2)} = \frac{v_1^{\frac{1}{6}} v_2^{\frac{2}{6}} v_3^{\frac{3}{6}} \times (\sigma_1^2)^{\frac{1}{6}} (\sigma_2^2)^{\frac{2}{6}} (\sigma_3^2)^{\frac{3}{6}}}{\frac{1}{6}\sigma_1^2 v_1 + \frac{2}{6}\sigma_2^2 v_2 + \frac{3}{6}\sigma_3^2 v_3}. \qquad \ldots\ldots(1)$$

The result depends only on the relative magnitude of the three values of σ^2. Six different cases were taken and for each the 500 values of L_1^{-1} were computed; since the same basic values of v_1, v_2 and v_3 were used, the six resulting frequency distributions of L_1^{-1} are not completely independent, but their relationship is of no simple character. The cases taken were as follows:

	σ_1^2	:	σ_2^2	:	σ_3^2
Case 1	2		1		2
„ 2	1		2		2
„ 3	2		1		1
„ 4	1		2		1
„ 5	4		2		1
„ 6	10		2		1

Since the three values of n are unequal, the first four cases correspond to different situations. The resulting histograms for the six distributions of L_1^{-1} are shown in Figs. 1 and 2.

For each case the following steps were then performed:

(i) The appropriate moments μ_1' and μ_2' and hence $\mu_2 = \mu_2' - (\mu_1')^2$ were calculated from Wilks's equations (6) and (7).

(ii) These moment values were inserted into his equation (12) to give the constants a and b.

(iii) These constants were inserted in turn into his Type III equation (11), the curves drawn and frequencies calculated to enable a χ^2 test to be applied.

A summary of results is shown in Table I. The column headed $P\{\chi^2 > \chi_0^2\}$ shows the result of applying the χ^2 test for goodness of fit, χ_0^2 being the observed value. The agreement between the theoretical curves and the experimental sampling results is very close, and suggests that the method of approximation to the unknown true distribution of L_1^{-1} is most satisfactory for practical purposes.

Of course the investigation only covers the case $k = 3$ and $n_1 = 5$, $n_2 = 10$, $n_3 = 15$, but at any rate for larger samples one would not expect worse agreement.

It is also of interest to investigate, in the six cases, what would be the chance of detecting from the samples that σ_1, σ_2 and σ_3 were not all equal. Suppose that we used a rule of rejecting the hypothesis, H_0, that $\sigma_1 = \sigma_2 = \sigma_3$ whenever L_1

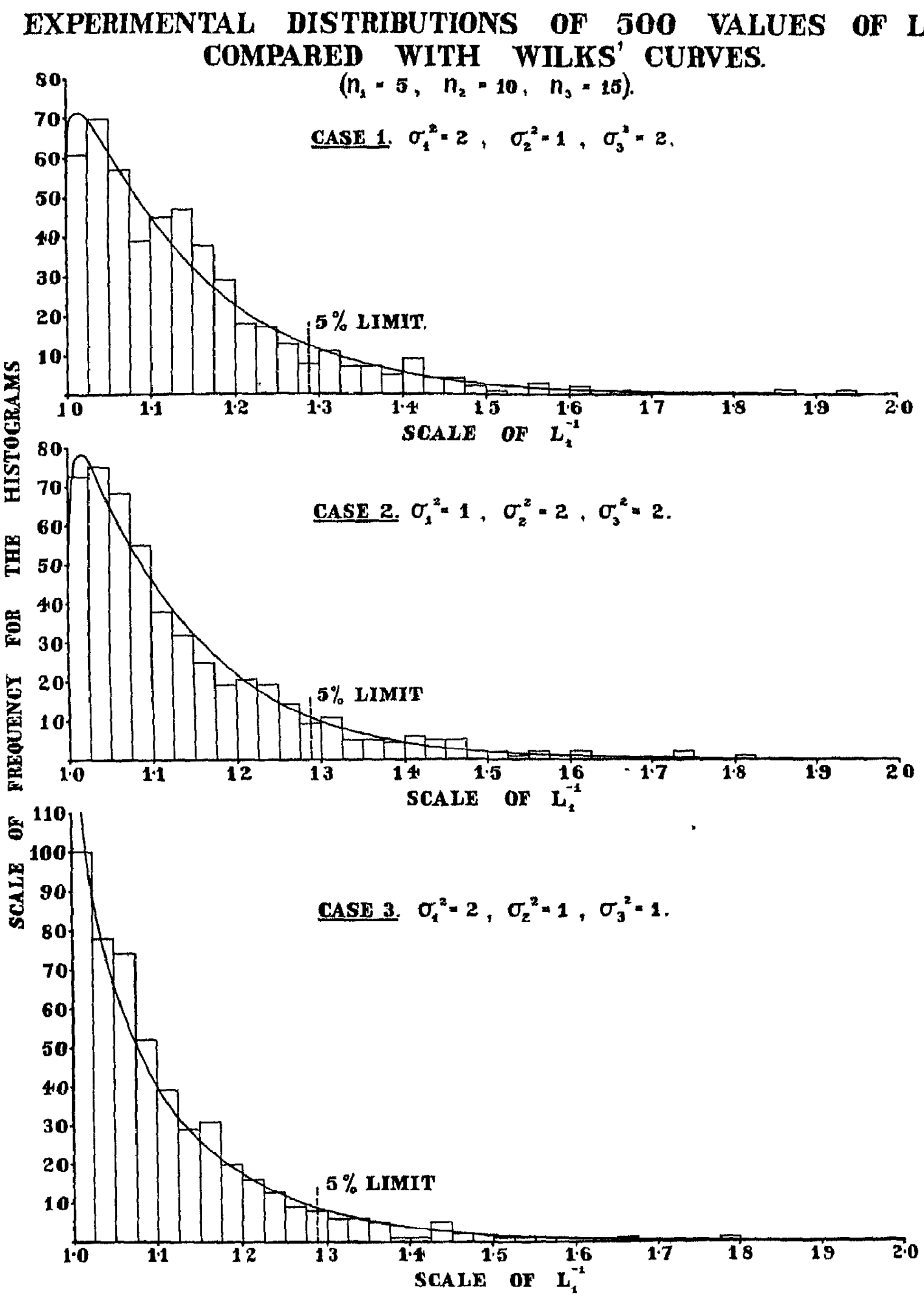


Fig. 1. N.B. The 5 % limit is appropriate for the case where the hypothesis tested is true.

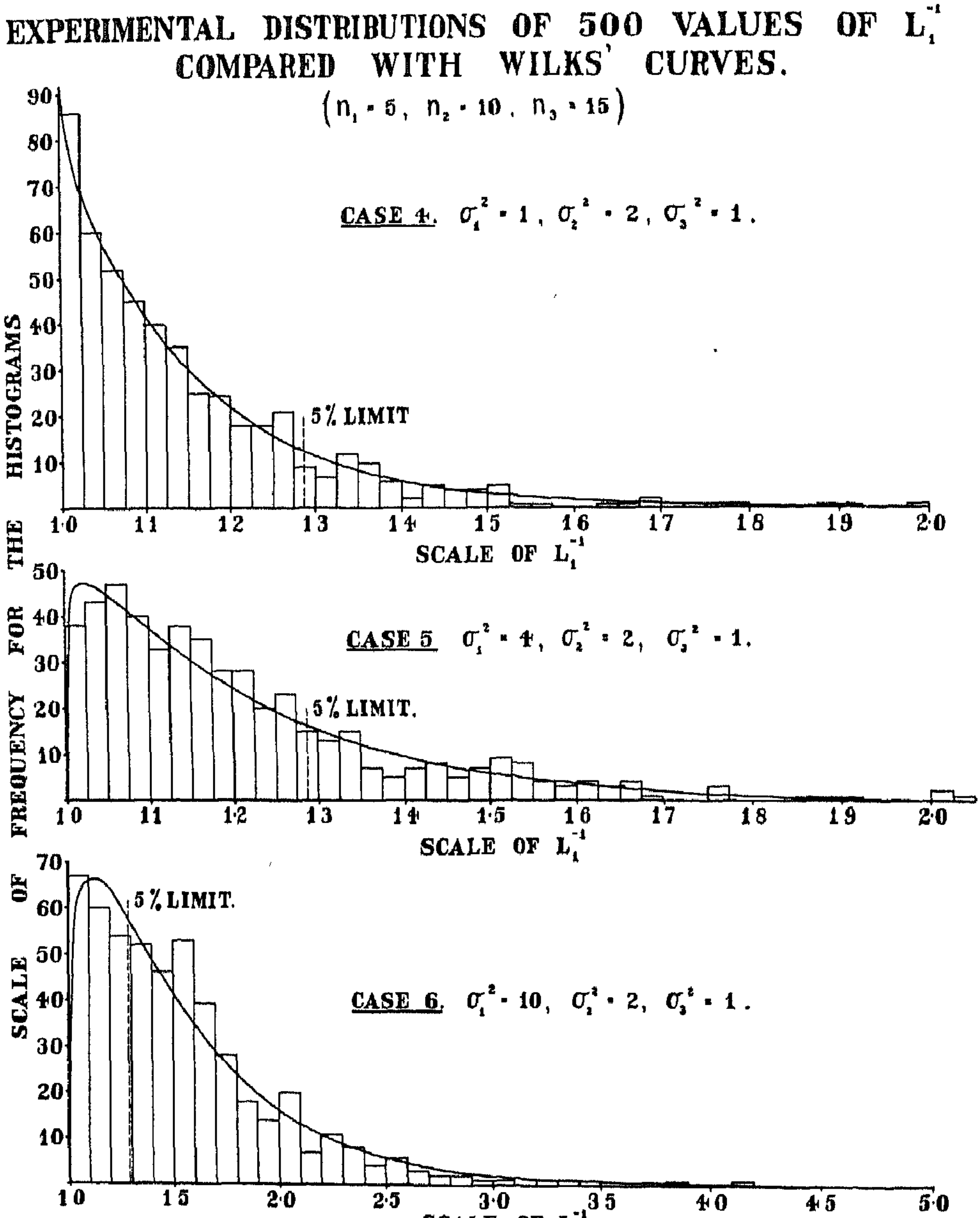


Fig. 2. N.B. The 5 % limit is appropriate for the case where the hypothesis tested is true.

TABLE I

Summary of results of calculations

Case	Values of the σ^2's	Moments	Theory	Observed	a	b	$P\{\chi^2 > \chi_0^2\}$	Power of test*
1	2 : 1 : 2	μ_1'	1·1462	1·1426	1·0891	7·4493	·615	·136
		μ_2	0·0196	0·0168				
2	1 : 2 : 2	μ_1'	1·1356	1·1325	1·0807	7·9725	·965	·120
		μ_2	0·0170	0·0160				
3	2 : 1 : 1	μ_1'	1·1152	1·1082	0·8197	7·1144	·525	·093
		μ_2	0·0162	0·0115				
4	1 : 2 : 1	μ_1'	1·1553	1·1495	0·9747	6·2760	·853	·158
		μ_2	0·0247	0 0230				
5	4 : 2 : 1	μ_1'	1·2195	1·2071	1·1013	5·0165	·816	·271
		μ_2	0·0438	0·0330				
6	10 : 2 : 1	μ_1'	1·5663	1·5283	1·2543	2·2148	·389	·646
		μ_2	0·2557	0·1968				

* Using 5 % significance level.

falls beyond the 5 % level, say $L_1\,(0{\cdot}05)$. It is first necessary to calculate this level for the particular case considered. To do this, Neyman & Pearson's Type I approximation* to the distribution of L_1 if H_0 is true, namely

$$p\,(L_1) = \frac{\Gamma\,(m_1+m_2)}{\Gamma\,(m_1)\,\Gamma\,(m_2)}\,L_1^{m_1-1}\,(1-L_1)^{m_2-1} \qquad \text{......(2)}$$

may be used. The true sampling moments of L_1 about zero are in this case

$$\left.\begin{aligned} \mu_1' &= \frac{2N}{N-3}\prod_{t=1}^{k}\left\{\frac{\Gamma\left(\dfrac{n_t-1}{2}+\dfrac{n_t}{N}\right)}{n_t^{\frac{n_t}{N}}\,\Gamma\left(\dfrac{n_t-1}{2}\right)}\right\} = 0{\cdot}9225\,2768,\\ \mu_2' &= \frac{4N^2}{(N-3)\,(N-1)}\prod_{t=1}^{3}\left\{\frac{\Gamma\left(\dfrac{n_t-1}{2}+\dfrac{2n_t}{N}\right)}{n_t^{\frac{2n_t}{N}}\,\Gamma\left(\dfrac{n_t-1}{2}\right)}\right\} = 0{\cdot}8561\,9647. \end{aligned}\right\} \qquad \text{......(3)}$$

* See reference on p. 124 above and also P. P. N. Nayer, *loc. cit.*

The values for m_1 and m_2 are then chosen so that the first two moments of the distribution (2) have the values (3), that is to say,

$$m_1 = \frac{\mu_1'(\mu_1'-\mu_2')}{\mu_2'-(\mu_1')^2} = 11{\cdot}90710, \quad m_2 = \frac{(1-\mu_1')(\mu_1'-\mu_2')}{\mu_2'-(\mu_1')^2} = 0{\cdot}99994. \quad \ldots\ldots(4)$$

In this case the value of m_2 is so close to unity that we may use the approximation

$$p(L_1) = m_1 L_1^{m_1-1}, \qquad \ldots\ldots(5)$$

and hence for the 5 % limit

$$0{\cdot}05 = \int_0^{L_1(0\cdot05)} p(L_1)\,dL_1 = \{L_1(0{\cdot}05)\}^{m_1}$$

and

$$L_1(0{\cdot}05) = 0{\cdot}778. \qquad \ldots\ldots(6)$$

This limit, or rather the limit $\{L_1(0{\cdot}05)\}^{-1} = 1{\cdot}286$, has been drawn in each of the diagrams. Neyman & Pearson* have defined the power of a test with regard to an alternative hypothesis H_1 as the probability that it will reject the hypothesis tested, H_0, when H_1 is true. Thus if a 5 % level of significance is used, the power of the L_1 test in the six cases illustrated is given by the proportionate area under Wilks's curves lying to the right of the critical levels drawn at $L_1^{-1} = 1{\cdot}286$. The six values of this probability, obtained by quadrature, are given in Table I.

It is interesting to note how, owing to the three samples being of unequal size, the power is different in each of the first four cases. Thus when the samples of 10 and 15 come from populations with the same variance, $\sigma_2^2 = \sigma_3^2$, and the smallest sample of 5 from a population with twice the variance (case 3), the test is least likely to detect the difference. It is somewhat more likely to do so when $\sigma_1^2 = \frac{1}{2}\sigma_2^2 = \frac{1}{2}\sigma_3^2$ (case 2). We also see how difficult it is to detect differences in population variances when only small samples are available; even in case 6, where $\sigma_1^2 : \sigma_2^2 : \sigma_3^2 = 10 : 2 : 1$, the odds are only 2 to 1 in favour of our being able to discover the difference, using the L_1 test.

* Neyman & Pearson, "Contribution to the Theory of testing Statistical Hypotheses," *Statistical Research Memoirs*, I (1936), pp. 1–37.

THE EXACT VALUE OF THE MOMENTS OF THE DISTRIBUTION OF χ^2, USED AS A TEST OF GOODNESS OF FIT, WHEN EXPECTATIONS ARE SMALL

By J. B. S. HALDANE, F.R.S.

1. Introduction to general method

In genetical practice we are constantly presented with large numbers of small samples from populations consisting of several well-defined classes. For example in the mouse we can readily obtain hundreds of litters containing anything from one up to about twelve members. Their totals may agree satisfactorily with expectation on a Mendelian basis, for example $\frac{1}{2}$ coloured, $\frac{1}{2}$ white, or $\frac{9}{16}$ grey, $\frac{3}{16}$ black, $\frac{1}{4}$ white. But we desire to know whether the individual litters can be regarded as random samples from such a population. In addition the problem of homogeneity may arise. That is to say the population as a whole may not conform to any particular expectation. But we may desire to know whether the litters can be regarded as random samples of the population given by the totals.

It has long been known that when the numbers expected in any observation are small, the distribution of χ^2 departs from that given by Pearson (1900). The mean appears sometimes, but not always, to be equal to the number of degrees of freedom. But the variance is no longer exactly equal to twice that number. Exact expressions for it in certain cases have been given by Pearson (1932) and Cochran (1936). These are based on an ingenious application of the theory of multiple contingency by Pearson.

It will be shown in this paper that the first few moments can often be calculated by entirely elementary methods involving nothing more advanced than the multinomial theorem. In an accompanying paper (Grüneberg and Haldane, 1937) they will be applied to actual data on mice.

We first study the distribution of χ^2 in a n-fold table with $n-1$ degrees of freedom, then in a $(m \times n)$-fold table with $m(n-1)$ degrees of freedom. For genetical work we are particularly interested in the $(n \times 2)$-fold table with n degrees of freedom. As a limiting case of the 2-fold table with 1 degree of freedom we derive the moments of the variance of samples from a Poisson series, and thence the distribution of χ^2 in a n-fold table with n degrees of freedom. The important case of the $(m \times n)$-fold table with $(m-1)$ $(n-1)$ degrees of freedom remains to be investigated.

Consider a sample of s individuals falling into n classes. Let the expected and observed numbers in these classes be:

$$\text{Expected } p_1 s, p_2 s, \ldots, p_i s, \ldots, p_n s,$$

$$\text{Observed } a_1, a_2, \ldots, a_i, \ldots, a_n,$$

where $\sum_{i=1}^{n} p_i = 1$, $\sum_{i=1}^{n} a_i = s$. It is assumed that we are sampling from an infinite population, or that if it is finite, any individual observed is replaced before the next individual is chosen at random. We shall use the notation $s_n = \frac{s!}{(s-n)!}$, and $E[x]$ to denote the expected value of x, or $\bar{x}$.

The probability of obtaining in a sample exactly a_1, a_2, a_3, etc., members of the n classes is $s! \prod_{i=1}^{n} \frac{p_i^{a_i}}{a_i!}$.

Hence the expected value of

$$\frac{a_f!}{(a_f-\alpha_f)!}\frac{a_g!}{(a_g-\alpha_g)!}\frac{a_h!}{(a_h-\alpha_h)!}$$

is the sum of this quantity multiplied by

$$s! \prod_{i=1}^{n} \frac{p_i^{a_i}}{a_i!},$$

summation being over all permissible sets of values of a_f, a_g and a_h, i.e. for all zero or positive integral values satisfying the condition $\sum_{i=1}^{n} (a_i) = s$.* Making use of the multinomial theorem it is seen that this sum is $p_f^{\alpha_f} p_g^{\alpha_g} p_h^{\alpha_h} s_{(\alpha_f+\alpha_g+\alpha_h)}$. But we can readily express any power of a_f, or any multiple of powers such as $a_f^4 a_g^2 a_h^2$, as a sum of expressions of the form $\frac{a_i!}{(a_i-\alpha_i)!}$ and of their products. Hence we can express the expected value of any power or product of powers as a sum of terms of the form $p_f^r p_g^s p_h^t s_{r+s+t}$. The following expressions for the expected values of powers and their products will be required in the analysis which follows:

$$E[a_i^2] = p_i^2 s_2 + p_i s,$$

$$E[a_i^4] = p_i^4 s_4 + 6p_i^3 s_3 + 7p_i^2 s_2 + p_i s,$$

$$E[a_i^6] = p_i^6 s_6 + 15p_i^5 s_5 + 65p_i^4 s_4 + 90p_i^3 s_3 + 31p_i^2 s_2 + p_i s,$$

$$E[a_i^8] = p_i^8 s_8 + 28p_i^7 s_7 + 266p_i^6 s_6 + 1050p_i^5 s_5 + 1701p_i^4 s_4 + 966p_i^3 s_3 + 127p_i^2 s_2 + p_i s.$$

In general $E[a_i^n] = \Sigma_r {}_rU_n p_i^r s_r$, where ${}_rU_n = {}_{r-1}U_{n-1} + r\,{}_rU_{n-1}$.†

$$E[a_i^2 a_j^2] = p_i^2 p_j^2 s_4 + (p_i^2 p_j + p_i p_j^2) s_3 + p_i p_j s_2,$$

$$E[a_i^4 a_j^2] = p_i^4 p_j^2 s_6 + (p_i^4 p_j + 6p_i^3 p_j^2) s_5 + (6p_i^3 p_j + 7p_i^2 p_j^2) s_4 + (7p_i^2 p_j + p_i p_j^2) s_3 + p_i p_j s_2,$$

$$E[a_i^2 a_j^2 a_k^2] = p_i^2 p_j^2 p_k^2 s_6 + (p_i p_j^2 p_k^2 + p_i^2 p_j p_k^2 + p_i^2 p_j^2 p_k) s_5 + (p_i^2 p_j p_k + p_i p_j^2 p_k + p_i p_j p_k^2) s_4 + p_i p_j p_k s_3,$$

* The value of $a_i!/(a_i-\alpha_i)! = a_i(a_i-1)\ldots(a_i-\alpha_i+1)$ is of course zero for $a_i < \alpha_i$.

† Mr C. Eisenhart has kindly pointed out to me that these coefficients are differences of powers of zero divided by appropriate factorials.

$$E[a_i^6a_j^2]=p_i^6p_j^2s_8+(p_i^6p_j+15p_i^5p_j^2)s_7+(15p_i^5p_j+65p_i^4p_j^2)s_6+(65p_i^4p_j+90p_i^3p_j^2)s_5$$
$$+(90p_i^3p_j+31p_i^2p_j^2)s_4+(31p_i^2p_j+p_ip_j^2)s_3+p_ip_js_2,$$
$$E[a_i^4a_j^4]=p_i^4p_j^4s_8+(6p_i^4p_j^3+6p_i^3p_j^4)s_7+(7p_i^4p_j^2+36p_i^3p_j^3+7p_i^2p_j^4)s_6$$
$$+(p_i^4p_j+42p_i^3p_j^2+42p_i^2p_j^3+p_ip_j^4)s_5+(6p_i^3p_j+49p_i^2p_j^2+6p_ip_j^3)s_4$$
$$+(7p_i^2p_j+7p_ip_j^2)s_3+p_ip_js_2,$$
$$E[a_i^4a_j^2a_k^2]=p_i^4p_j^2p_k^2s_8+(p_i^4p_j^2p_k+p_i^4p_jp_k^2+6p_i^3p_j^2p_k^2)s_7$$
$$+(p_i^4p_jp_k+6p_i^3p_j^2p_k+6p_i^3p_jp_k^2+7p_i^2p_j^2p_k^2)s_6$$
$$+(6p_i^3p_jp_ks_5+7p_i^2p_j^2p_k+7p_i^2p_jp_k^2+p_ip_j^2p_k^2)s_5$$
$$+(7p_i^2p_jp_k+p_ip_j^2p_k+p_i^2p_jp_k^2)s_4+p_ip_jp_ks_3,$$
$$E[a_i^2a_j^2a_k^2a_l^2]=p_i^2p_j^2p_k^2p_l^2s_8+\Sigma p_i^2p_j^2p_k^2p_ls_7+\Sigma p_i^2p_j^2p_kp_ls_6$$
$$+\Sigma p_i^2p_jp_kp_ls_5+p_ip_jp_kp_ls_4.$$

In the special case where $p_1=p_2=p_3=\ldots=n^{-1}$ we have such expressions as

$$E[a_i^4a_j^2a_k^2]=n^{-8}s_8+8n^{-7}s_7+20n^{-6}s_6+21n^{-5}s_5+9n^{-4}s_4+s_3.$$

In what follows, Σ denotes summation over all values of i, $\Sigma\Sigma$ summation over all pairs of unequal values of i and j. The following notation is used for sums of reciprocals: $R_1=\Sigma p_i^{-1}$, $R_2=\Sigma p_i^{-2}$, $R_3=\Sigma p_i^{-3}$.

$$\chi^2=\Sigma\frac{(p_is-a_i)^2}{p_is}$$
$$=\frac{1}{s}\Sigma\frac{a_i^2}{p_i}-s.$$

Hence
$$\overline{\chi^2}=s^{-1}(s_2\Sigma p_i+s\Sigma 1)-s$$
$$=n-1.$$

$$\overline{\chi^4}=E\left[s^{-2}\Sigma\frac{a_i^4}{p_i^2}+2s^{-2}\Sigma\Sigma\frac{a_i^2a_j^2}{p_ip_j}-2\Sigma\frac{a_i^2}{p_i}+s^2\right]$$
$$=s^{-2}[s_4(\Sigma p_i^2+2\Sigma\Sigma p_ip_j)+s_3(6\Sigma p_i+4\Sigma\Sigma p_i)+s_2(7\Sigma_1+2\Sigma\Sigma 1)+s\Sigma p_i^{-1}]$$
$$-2(s_2\Sigma p_i+s\Sigma 1)+s^2$$
$$=s^{-2}[s_4+(2n+4)s_3+(n^2+6n)s_2+R_1s]-2(s_2+ns)+s^2$$
$$=n^2-1+(R_1-n^2-2n+2)s^{-1}.$$

Hence
$$\mu_2=2(n-1)+(R_1-n^2-2n+2)s^{-1}.$$

This agrees with Pearson's (1932) result. The calculation of the higher moments is somewhat tedious. It can be greatly simplified by the following device. The moments are calculated for the special case when all p_i's are equal. The terms in the general case which involve sums of negative powers of the p_i's are then calculated separately, and an adjustment made to the previous formulae, since when $p_i=n^{-1}$, $R_k=n^{k+1}$.

2. Case of an n-fold classification ($n-1$ degrees of freedom)

For a sample of s divided into n classes of which the expectations are equal, we find:

$$E[\Sigma a_i^2] = n^{-1}(s_2 + ns) = n^{-1}[s^2 + (n-1)s],$$

$$\begin{aligned} E[\Sigma a_i^2]^2 &= n^{-2}[s_4 + (2n+4)s_3 + (n^2+6n)s_2 + n^2 s] \\ &= n^{-2}[s^4 + 2(n-1)s^3 + (n^2-1)s^2 - 2(n-1)s], \end{aligned}$$

$$\begin{aligned} E[\Sigma a_i^2]^3 &= n^{-3}[s_6 + (3n+12)s_5 + (3n^2+30n+32)s_4 + (n^3+21n^2+68n)s_3 \\ &\qquad + (3n^3+28n^2)s_2 + n^3 s] \\ &= n^{-3}[s^6 + 3(n-1)s^5 + 3(n^2-1)s^4 + (n^3+3n^2-7n+3)s^3 \\ &\qquad - 2(n^2+12n-13)s^2 - 4(n^2-7n+6)s], \end{aligned}$$

$$\begin{aligned} E[\Sigma a_i^2]^4 &= n^{-4}[s_8 + (4n+24)s_7 + (6n^2+84n+176)s_6 + (4n^3+102n^2+544n+400)s_5 \\ &\quad + (n^4+48n^3+516n^2+1136n)s_4 + (6n^4+152n^3+808n^2)s_3 + (7n^4+120n^3)s_2 + n^4 s] \\ &= n^{-4}[s^8 + 4(n-1)s^7 + 6(n^2-1)s^6 + 4(n^3+3n^2-4n)s^5 \\ &\qquad + (n^4+8n^3+6n^2-104n+89)s^4 + 4(n^3-17n^2-45n+61)s^3 \\ &\qquad - 4(2n^3+51n^2-314n+261)s^2 - 8(n^3-31n^2+120n-90)s]. \end{aligned}$$

Hence

$$\overline{\chi^2} = n-1,$$

$$\overline{\chi^4} = n^2 - 1 - 2(n-1)s^{-1},$$

$$\overline{\chi^6} = (n+3)(n+1)(n-1) - 2(n-1)(n+13)s^{-1} - 4(n-1)(n-6)s^{-2},$$

$$\begin{aligned} \overline{\chi^8} &= (n+5)(n+3)(n+1)(n-1) + 4(n-1)(n^2-12n-85)s^{-1} \\ &\qquad - 4(n-1)(2n^2+53n-261)s^{-2} - 8(n-1)(n^2-30n+90)s^{-3}. \end{aligned}$$

The first four moments and cumulants are:

$$\left.\begin{aligned} \mu_1' &= \kappa_1 = n-1, \\ \mu_2 &= \kappa_2 = 2(n-1) - 2(n-1)s^{-1} = 2s^{-1}(n-1)(s-1), \\ \mu_3 &= \kappa_3 = 8(n-1) + 4(n-1)(n-8)s^{-1} - 4(n-1)(n-6)s^{-2} \\ &\qquad = 4s^{-2}(n-1)(s-1)(n+2s-6), \\ \mu_4 &= 12(n-1)(n+3) + 24(n-1)(3n-19)s^{-1} + 4(n-1)(2n^2-81n+285)s^{-2} \\ &\qquad - 8(n-1)(n^2-30n+90)s^{-3}, \\ \kappa_4 &= 48(n-1) + 96(n-1)(n-5)s^{-1} + 8(n-1)(n^2-42n+144)s^{-2} \\ &\qquad - 8(n-1)(n^2-30n+90)s^{-3} \\ &= 8s^{-2}(n-1)(s-1)[n^2 + 6(2s-5)n + 6(s^2-9s+15)], \end{aligned}\right\} \quad \ldots\ldots(1)$$

where $\kappa_4 = \mu_4 - 3\mu_2^2$.

Hence $$\beta_1 = \frac{2(n+2s-6)^2}{(n-1)s(s-1)}.$$

In the general case where not all the p_i's are equal, we find:

$$E\left[\Sigma\frac{a_i^2}{p_i}\right] = s_2 + ns = s^2 + (n-1)s,$$

$$E\left[\Sigma\frac{a_i^2}{p_i}\right]^2 = s_4 + (2n+4)s_3 + (n^2+6n)s_2 + R_1 s$$
$$= s^4 + 2(n-1)s^3 + (n^2-1)s^2 + (R_1 - n^2 - 2n + 2)s,$$

$$E\left[\Sigma\frac{a_i^2}{p_i}\right]^3 = s_6 + (3n+12)s_5 + (3n^2+30n+32)s_4 + (3R_1 + n^3 + 18n^2 + 68n)s_3 + (3n+28)R_1 s_2 + R_1 s$$
$$= s^6 + 3(n-1)s^5 + 3(n^2-1)s^4 + (3R_1 + n^3 - 7n + 3)s^3 + [(3n+19)R_1 - 3n^3 - 21n^2 - 24n + 26]s^2 + [R_2 - (3n+22)R_1 + 2n^3 + 18n^2 + 28n - 24]s,$$

$$E\left[\Sigma\frac{a_i^2}{p_i}\right]^4 = s_8 + (4n+24)s_7 + (6n^2+84n+176)s_6 + (6R_1 + 4n^3 + 96n^2 + 544n + 400)s_5 + [(12n+36)R_1 + n^4 + 36n^3 + 380n^2 + 1136n]s_4 + [4R_2 + (6n^2+148n+808)R_1]s_3 + [(4n+120)R_2 + 3R_1^2]s_2 + R_3 s$$
$$= s^8 + 4(n-1)s^7 + 6(n^2-1)s^6 + (6R_1 + 4n^3 + 6n^2 - 16n)s^5 + [4(3n+19)R_1 + n^4 - 4n^3 - 70n^2 - 104n + 89]s^4 + [4R_2 + (6n^2+76n+202)R_1 - 6n^4 - 76n^3 - 270n^2 - 180n + 244]s^3 + [(4n+108)R_2 + 3R_1^2 - (18n^2+312n+1228)R_1 + 11n^4 + 196n^3 + 1024n^2 + 1256n - 1044]s^2 + [R_3 - (4n+112)R_2 - 3R_1^2 + (12n^2+224n+944)R_1 - 6n^4 - 120n^3 - 696n^2 - 960n + 720]s.$$

But $$\chi^{2n} = \left(s^{-1}\Sigma\frac{a_i^2}{p_i} - s\right)^n.$$

Hence

$$\overline{\chi^2} = n-1,$$

$$\overline{\chi^4} = (n+1)(n-1) + (R_1 - n^2 - 2n + 2)s^{-1},$$

$$\overline{\chi^6} = (n+3)(n+1)(n-1) + [(3n+19)R_1 - (3n^3 + 21n^2 + 24n - 26)]s^{-1} + [R_2 - (3n+22)R_1 + 2n^3 + 18n^2 + 28n - 24]s^{-2},$$

$$\overline{\chi^8} = (n+5)(n+3)(n+1)(n-1) + 2[(3n^2+44n+145)R_1 - (32n^4 + 42n^3 + 171n^2 + 146n - 170)]s^{-1} + [4(n+27)R_2 + 3R_1^2 - 2(9n^2+156n+614)R_1 + 11n^4 + 196n^3 + 1024n^2 + 1256n - 1044]s^{-2} + [R_3 - 4(n+28)R_2 - 3R^2 + 4(3n^2+56n+236)R_1 - 6(n^4 + 20n^3 + 116n^2 + 160n - 120)]s^{-3}.$$

Therefore,

$$\left.\begin{aligned}
&\mu_1' = \kappa_1 = n-1,\\
&\mu_2 = \kappa_2 = 2(n-1) + [R_1 - (n^2+2n-2)]\,s^{-1},\\
&\mu_3 = \kappa_3 = 8(n-1) + 2[11R_1 - (9n^2+18n-16)]\,s^{-1}\\
&\qquad\qquad\qquad + [R_2 - (3n+22)R_1 + 2(n^3+9n^2+14n-12)]\,s^{-2},\\
&\mu_4 = 12(n-1)(n+3) + 12[(n+31)R_1 - (n^3+25n^2+44n-38)]\,s^{-1}\\
&\quad + [112R_2 + 3R_1^2 - 2(3n^2+118n+658)R_1 + 3(n^4+44n^3+328n^2+488n-380)]\,s^{-2}\\
&\quad + [R_3 - 4(n+28)R_2 - 3R_1^2 + 4(3n^2+56n+236)R_1\\
&\qquad\qquad\qquad - 6(n^4+20n^3+116n^2+160n-120]\,s^{-3},\\
&\kappa_4 = 48(n-1) + 96(4R_1 - 3n^2 - 6n + 5)\,s^{-1}\\
&\quad + 8[14R_2 - 2(14n+83)R_1 + 3(5n^3+41n^2+62n-48)]\,s^{-2}\\
&\quad + [R_3 - 4(n+28)R_2 - 3R_1^2 + 4(3n^2+56n+236)R_1\\
&\qquad\qquad\qquad - 6(n^4+20n^3+116n^2+160n-120)]\,s^{-3}.
\end{aligned}\right\} \quad \ldots\ldots(2)$$

It will be seen that when any of the expected frequencies p_i is very small, the moments may be considerably larger than those of the classical χ^2, to which they approximate when the number s in the sample is large.

3. Special case of two classes (1 degree of freedom)

When there are only two expected classes, with frequencies p and q, i.e. $n=2$, we have, if $k = \dfrac{1}{pq}$,

$$\left.\begin{aligned}
&\mu_1' = \kappa_1 = 1,\\
&\mu_2 = \kappa_2 = 2 + (k-6)\,s^{-1},\\
&\mu_3 = \kappa_3 = 8 + 2(11k-6)\,s^{-1} + (k^2 - 30k + 120)\,s^{-2},\\
&\mu_4 = 60 + 12(33k-158)\,s^{-1} + (115k^2 - 2036k + 6828)\,s^{-2}\\
&\qquad\qquad\qquad + (k^3 - 126k^2 + 1680k - 5040)\,s^{-3},\\
&\kappa_4 = 48 + 96(4k-19)\,s^{-1} + 16(7k^2 - 125k + 420)\,s^{-2}\\
&\qquad\qquad\qquad + (k^3 - 126k^2 + 1680k - 5040)\,s^{-3}.
\end{aligned}\right\} \quad \ldots\ldots(3)$$

These expressions can of course be calculated independently, and furnish a useful check on the equations (2). When s tends to infinity, provided neither p nor q tends to zero, we have $\kappa_2 = 2$, $\kappa_3 = 8$, $\kappa_4 = 48$, $\kappa_n = 2^{n-1}(n-1)\,!$, the values appropriate to Pearson's χ^2. If, however, sp remains equal to g while s tends to infinity we have:

$$\left.\begin{aligned}
&\mu_1' = \kappa_1 = 1,\\
&\mu_2 = \kappa_2 = 2 + g^{-1},\\
&\mu_3 = \kappa_3 = 8 + 22g^{-1} + g^{-2},\\
&\mu_4 = 60 + 396g^{-1} + 115g^{-2} + g^{-3},\\
&\kappa_4 = 48 + 384g^{-1} + 112g^{-2} + g^{-3}.
\end{aligned}\right\} \quad \ldots\ldots(4)$$

These are the moments and cumulants of $\chi^2=(a-g)^2/g$ for a sample from a Poisson series when the expected value is g and the observed value is a. If V be the variance, $V=(a-g)^2=g\chi^2$, so $\mu_n g^n$ and $\kappa_n g^n$ are the moments and cumulants of the variance of such a sample.

4. Case of $(m\times n)$-fold classification ($m(n-1)$ degrees of freedom)

We now consider the values of χ^2 in 2 dimensional tables. Consider a $(m\times n)$-fold table where m samples of $s_1, s_2, s_3, \ldots, s_r, \ldots, s_m$ members have been drawn independently from an infinite population in which the frequencies of n classes are $p_1, p_2, p_3, \ldots, p_i, \ldots, p_n$, and where, as above,

$$R_1=\sum_{i=1}^{n} p_i^{-1},\quad R_2=\sum_{i=1}^{n} p_i^{-2},\quad R_3=\sum_{i=1}^{n} p_i^{-3}.$$

There are clearly $m(n-1)$ degrees of freedom. Summing the cumulants, given in equation (2), appropriate to the χ^2 calculated from each sample, we have therefore:

$$\left.\begin{aligned}
\mu_1' &= \kappa_1 = m(n-1),\\
\mu_2 &= \kappa_2 = 2m(n-1)+[R_1-(n^2+2n-2)]\sum_{r=1}^{m} s_r^{-1},\\
\mu_3 &= \kappa_3 = 8m(n-1)+2[11R_1-(9n^2+18n-16)]\sum_{r=1}^{m} s_r^{-1}\\
&\qquad +[R_2-(3n+22)R_1+2(n^3+9n^2+14n-12)]\sum_{r=1}^{m} s_r^{-2},\\
\kappa_4 &= 48m(n-1)+96[4R_1-(3n^2+6n-5)]\sum_{r=1}^{m} s_r^{-1}\\
&\qquad +8[14R_2-2(14n+83)R_1+3(5n^3+41n^2+62n-48)]\sum_{r=1}^{m} s_r^{-2}\\
&\qquad +[R_3-4(n+28)R_2-3R_1^2+4(3n^2+56n+236)R_1\\
&\qquad\qquad -6(n^4+20n^3+116n^2+160n-120)]\sum_{r=1}^{m} s_r^{-3},\\
\mu_4 &= \kappa_4+3\kappa_2^2.
\end{aligned}\right\}\qquad\ldots\ldots(5)$$

5. Case of $(n\times 2)$-fold classification (n degrees of freedom)

These are the most general formulae arrived at in this paper. Special cases analogous to equations (1) and (3) obviously arise. Only the latter will be given as it is used by Grüneberg and Haldane (1937). For n samples each consisting of s members, and each divided into two classes, whose expected values are ps and qs, we have

$$\chi^2=\sum_{r=1}^{n}\frac{(a_r-ps)^2}{spq},\ \text{degrees of freedom}=n,$$

where a_r is observed frequency in the first class of rth sample. Further, writing $k=(pq)^{-1}$ we have for the moments and cumulants of χ^2:

$$\mu_1'=\kappa_1=n,$$
$$\mu_2=\kappa_2=ns^{-1}(2s+k-6),$$
$$\mu_3=\kappa_3=ns^{-2}[8s^2+(22k-112)s+k^2-30k+120],$$
$$\mu_4=ns^{-3}[12(n+4)s^3+12\{(k-6)n+8(4k-19)\}s^2 \\ +\{3n(k-6)^2+16(7k^2-125k+420)\}+k^3-126k^2+1680k-5040],$$
$$\kappa_4=ns^{-3}[48s^3+96(4k-19)s^2+16(7k^2-125k+420)s \\ +k^3-126k^2+1680k-5040]. \qquad \ldots\ldots(6)$$

Hence $$\beta_1=\frac{[8s^2+(22k-112)s+k^2-30k+120]^2}{ns(2s+k-6)^3}.$$

It will be seen that when k is large compared with s, that is to say when one of the expectations is a small fraction, β_1 approximates to k/ns, whereas its value when s tends to infinity is $8/n$. But for moderate values of s the skewness may be considerably less than in the classical case.

Two numerical values of k are important in genetics. If $p=q=\frac{1}{2}$, $k=4$, and:

$$\left.\begin{aligned} \mu_2=\kappa_2&=2n(s-1)s^{-1},\\ \mu_3=\kappa_3&=8n(s-1)(s-2)s^{-2},\\ \kappa_4&=16n(s-1)(3s^2-15s+17)s^{-3},\\ \beta_1&=\frac{8(s-2)^2}{ns(s-1)}. \end{aligned}\right\} \qquad \ldots\ldots(7)$$

Thus for example when $n=50$, if $s=4$, $\beta_1=\cdot053$, whereas when s is infinite $\beta_1=\cdot16$.

When $p=\frac{1}{4}$, $q=\frac{3}{4}$, $k=\frac{16}{3}$, we have:

$$\left.\begin{aligned} \mu_2=\kappa_2&=\frac{2n(3s-1)}{3s},\\ \mu_3=\kappa_3&=\frac{8n(9s^2+6s-13)}{9s^2},\\ \kappa_4&=\frac{16n(81s^3+378s^2-1284s+823)}{27s^2},\\ \beta_1&=\frac{8(9s^2+6s-13)^2}{3ns(3s-1)^3}. \end{aligned}\right\} \qquad \ldots\ldots(8)$$

For example if $n=50$, $s=4$, $\beta_1=\cdot24$.

The values of s will generally vary from one sample to another. In this case the values of the cumulants are the sums of the values found for the different sample sizes.

6. Case of n-fold classification (n degrees of freedom)

A limiting case of the $(n \times 2)$-fold classification arises when the values of s_r tend to infinity, but the expectations, $g_r = ps_r$, in say the first category remain finite. If a_r is the observed frequency in this category, it may be regarded as resulting from a single sample from a Poisson series with expected value g_r. Then

$$\chi^2 = \sum_{r=1}^{n} \frac{(a_r - g_r)^2}{g_r},$$

and using equations (4) we have

$$\begin{aligned}
&\overline{\chi^2} = n,\\
&\mu_2 = \kappa_2 = 2n + \Sigma g_r^{-1},\\
&\mu_3 = \kappa_3 = 8n + 22\,\Sigma g_r^{-1} + \Sigma g_r^{-2},\\
&\kappa_4 = 48n + 384\,\Sigma g_r^{-1} + 112\,\Sigma g_r^{-2} + \Sigma g_r^{-3}.
\end{aligned}$$

If we call s the size of the whole sample, and let $g_r = sp_r$, while $R_1 = \Sigma p_r^{-}$, $R_2 = \Sigma p_r^{-2}$, $R_3 = \Sigma p_r^{-3}$, we can write, in full analogy with equations (2):

$$\left.\begin{aligned}
&\mu_1' = \kappa_1 = n,\\
&\mu_2 = \kappa_2 = 2n + s^{-1}R_1,\\
&\mu_3 = \kappa_3 = 8n + 22s^{-1}R_1 + s^{-2}R_2,\\
&\mu_4 = 12n\,(n+4) + 12\,(n+32)\,s^{-1}R_1 + s^{-2}\,(3R_1^2 + 112R_2) + s^{-3}R_3,\\
&\kappa_4 = 48n + 384s^{-1}R_1 + 112s^{-2}R_2 + s^{-3}R_3.
\end{aligned}\right\} \quad \ldots(9)$$

The great simplicity of these expressions as compared with (2) is noteworthy. The extra terms in (2) represent diminutions in the moments due to the loss of one degree of freedom.

7. A weighting correction

If we have a number of samples of different sizes s_r their variances will differ. Now when s is large the probability that each sample will make a given contribution to χ^2 is equal. If we wish to reinstate this condition as far as possible, we must arrange for a proper weighting of the contributions made from the various samples.

If we have m samples, in each of which there are n classes in expected numbers sp_1, sp_2, etc., s being the number in the sample, taking the mean and variance from equation (2), we put for each sample:

$$\zeta = \frac{s^{\frac{1}{2}}[\chi^2 - (n-1)]}{[2\,(n-1)\,s - (n^2 + 2n - 2) + \Sigma p_i^{-1}]^{\frac{1}{2}}}.$$

Then in each sample the mean of ζ is 0, and its variance 1. Hence for the m

samples the variance of the sum of the ζ's is 0 and its variance n. In the case of a $(n \times 2)$-fold table with n degrees of freedom

$$\zeta = \frac{\chi^2 - 1}{\left(2 + \frac{k-6}{s}\right)^{\frac{1}{2}}}.$$

This weighting correction is hardly worth making in the Mendelian cases where $p = \frac{1}{2}$, $k = 4$, and $p = \frac{1}{4}$, $k = \frac{16}{3}$. Here the weighting factors vary from ·7071 for $s = \infty$ to 1 for $s = 2$ in the case when $k = 4$, and from ·7071 when $s = \infty$ to ·866 when $s = 1$ in the case when $k = \frac{16}{3}$. But when k is larger the variation is considerable. Thus, in the case of a selfed autotetraploid $p = \frac{1}{36}$, and the weighting factor falls from ·7071 when s is infinite to ·1740 when $s = 1$.

8. Discussion

The results obtained for the moments of χ^2 in a $(n \times 2)$-fold table when p is fixed do not agree with those given by Cochran (1936), who finds a mean value $n - 1$, and a variance $2\,(n-1) + \frac{(n-1)^2 (k-6)}{ns}$. These results only differ from those here given by a factor of the order $1 - \frac{1}{n}$, and are therefore satisfactory when n is large. However, when p is known and does not have to be estimated from the data there are clearly n degrees of freedom, and not $n - 1$; hence my own results would appear to be slightly more accurate than Cochran's. I have, however, no reason to doubt the accuracy of Cochran's results when p is estimated from the totals. It is noteworthy that while in the cases considered here, where χ^2 is used as a test of goodness of fit, its mean is always exactly equal to the number of degrees of freedom, this is no longer so when it is used as a test of homogeneity. Thus Cochran finds for a $(n \times 2)$-fold table with $n - 1$ degrees of freedom, a mean χ^2 of $n - 1 + \frac{n-1}{ns} + \frac{1}{ns^2}$, when all samples contain s members.

It follows from the results here given that the distribution of χ^2 for large values of n generally approximates fairly closely to normality. It would of course be possible to find a function of χ^2 whose distribution is much more nearly normal. Thus Wilson and Hilferty (1931) found that, when s is large,

$$\left[\left(\frac{\chi^2}{n}\right)^{\frac{1}{3}} - 1 + \frac{2}{9n}\right]\left(\frac{9n}{2}\right)^{\frac{1}{2}}$$

is very nearly normally distributed with mean zero and variance unity. It may also be desirable, as Fisher (1922), Neyman and Pearson (1928) and Cochran (1936) point out, to use the logarithm of the likelihood of the sample, rather than χ^2, as a test of goodness of fit when expectations are small. It is, however, worth

pointing out that, in the estimation of the frequency of lethal genes in autosomes, a problem with which I hope to deal later, χ^2 appears to furnish a simple and satisfactory estimate, and its distribution must therefore be known.

9. SUMMARY

Exact expressions have been found for the mean and the first four moments of χ^2 in cases where it is used as a test of goodness of fit, that is to say in $(m \times n)$-fold tables with $m(n-1)$ degrees of freedom. The mean is always exactly $m(n-1)$. The expressions for the higher moments are more complicated. Information has therefore been obtained which will make it possible to apply the χ^2 test without restriction on the size of the samples on the numbers expected. The results do not apply where χ^2 is used as a test of homogeneity, the expectations being deduced from observed totals.

REFERENCES

COCHRAN, W. G. (1936). "The χ^2 distribution for the binomial and Poisson series, with small expectations." *Ann. Eugen.*, Camb., VII, 207.

FISHER, R. A. (1922). "On the mathematical foundations of theoretical statistics." *Philos. Trans.* A, CCXXII, 309–68.

GRÜNEBERG, H. and HALDANE, J. B. S. (1937). "Test of goodness of fit applied to records of Mendelian segregation in mice." *Biometrika*, XXIX, 144–53.

NEYMAN, J. and PEARSON, E. S. (1928). "On the use and interpretation of certain test criteria for purposes of statistical inference. Part II." *Biometrika*, XX A, 263–94.

PEARSON, K. (1900). "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling." *Phil. Mag.* L, 157–75.

PEARSON, K. (1932). "Experimental Discussion of the (χ^2, P) test for goodness of fit." *Biometrika*, XXIV, 351–81.

WILSON and HILFERTY (1931). *Nat. Acad. Sci.* XVII, No 12, 684.

TESTS OF GOODNESS OF FIT APPLIED TO RECORDS OF MENDELIAN SEGREGATION IN MICE

By HANS GRÜNEBERG and J. B. S. HALDANE

Department of Genetics, University College, London

It has long been known that in some, though not in all instances, the totals in cases of segregation involving a small number of genes were in satisfactory accord with the simple numerical ratios expected according to Mendel's laws. Divergences could often be explained by selective mortality of zygotes between the time of fertilization and the time when the characters in question could be determined.

But where the totals were in agreement with Mendelian expectation it was not clear that individual families might not show an unexpected number, either larger or smaller than that expected on sampling theory, of large deviations from the expected ratio.

Where the families, and the expectations in all groups, were sufficiently large, it was possible to apply Pearson's classical χ^2 method (e.g. de Winton & Haldane, 1933). Where samples were smaller this was no longer possible. However, Haldane (1937, pp. 133–43 above) has calculated the moments of the distribution of χ^2 when expectations are small.

If we are dealing with n samples from a population in which two classes occur with frequency p and $1-p$, if s_r be the size of the rth sample, then the principal parameters of the distribution of χ^2 are given in Table I for the two Mendelian cases where $p = \frac{1}{2}$ and $\frac{1}{4}$. These results follow if we put $m = n$, $n = 2$ in Haldane's equation (5) (p. 139 above).

TABLE I

Distribution parameters of χ^2 in Mendelian cases

	$p=\frac{1}{2}$	$p=\frac{1}{4}$
Mean	n	n
Variance	$2(n-\Sigma s_r^{-1})$	$\frac{2}{3}(3n-\Sigma s_r^{-1})$
$\beta_1=\gamma_1^2$	$\dfrac{8(n-3\Sigma s_r^{-1}+2\Sigma s_r^{-2})^2}{(n-\Sigma s_r^{-1})^3}$	$\dfrac{8(9n+6\Sigma s_r^{-1}-13\Sigma s_r^{-2})^2}{3(3n-\Sigma s_r^{-1})^3}$
$\gamma_2=\beta_2-3$	$\dfrac{4(3n-18\Sigma s_r^{-1}+32\Sigma s_r^{-2}-17\Sigma s_r^{-3})}{(n-\Sigma s_r^{-1})^2}$	$\dfrac{4(81n+378\Sigma s_r^{-1}-1284\Sigma s_r^{-2}+823\Sigma s_r^{-3})}{3(3n-\Sigma s_r^{-1})^2}$

In the case of our data it will be shown that n is often so large that β_1 (or γ_1) and γ_2 are of the order of their own sampling errors or less. The distribution of χ^2 can thus be treated as approximately normal.

Our experimental data are as follows:

(1) Records of 562 litters including 3707 mice, from back-crosses of mice heterozygous for the normal colour gene **C** and its allelomorph $\mathbf{c^d}$ to recessives. These were obtained during an experiment (Gruneberg, 1936) on the linkage of **C** with two other genes. The ratios obtained for these two genes are closely correlated with those for **C** and $\mathbf{c^d}$, and hence will not be given, as they do not give independent information.

(2) Records of 273 litters including 1366 mice, from a back-cross of $\mathbf{Cc^d} \times \mathbf{c^dc^d}$ and reciprocally. These are selected. In many cases the full-coloured parent was not known to be heterozygous for $\mathbf{c^d}$. Families of less than 7 were rejected if the full-coloured parent was not known to be heterozygous. (By a family is meant a group of litters from the same two parents.)

(3) Records of 243 litters including 1198 mice, from matings $\mathbf{Cc^d} \times \mathbf{Cc^d}$. Here again families of less than 16 individuals were rejected unless both parents were known to be heterozygous.

(4) Records of 226 litters including 1279 mice, obtained by Fisher & Mather (1936) in the course of a linkage experiment, and very kindly put at our disposal by the authors. These litters were derived from matings of mice heterozygous for five or six genes, with multiple recessives. One of these genes (for recessive light head) was not recorded in all litters, and gave aberrant ratios. The other five, and sex, were recorded in all litters (except blue dilution in one). The records given to us cover 50 mice beyond those on which Fisher & Mather's published results are based.

All these data, totalling 7550 mice in 1304 litters, are collected in Table II. The first 562 litters are divided into five groups (I, II, III, IV and V) representing different experiments (see Gruneberg, 1936). The total is also given. The table is to be read as follows. The first column gives litter size. The second two, headed **D** and **r**, give distribution as between dominants and recessives, or in the case of (4) for sex, as between males and females. Subsequent columns give the numbers of litters of this type in the various experiments.

Table III gives the totals in the various experiments, each χ^2 having 1 degree of freedom. In only two cases (1, III and 4, blue) does the deviation exceed twice the standard error. If the data of Exp. 1 are combined and a single χ^2 found for them, the total χ^2 for the four experiments is 10·87 for 9 degrees of freedom, a very moderate value. If they are considered separately, $\chi^2 = 21{\cdot}86$ for 13 degrees of freedom. P now just exceeds 0·05, which is generally taken as the criterion of significance.

TABLE II

Litter size	D	r	1						2	3	4					
			I	II	III	IV	V	Total			Sex	a	wv	s	b	d
1	1	0	1	2	—	1	1	5	4	11	2	—	2	2	2	2
	0	1	—	4	—	2	—	6	7	3	2	4	2	2	2	2
2	2	0	—	—	—	—	1	1	7	10	1	3	3	2	3	1
	1	1	5	2	4	—	3	14	16	9	4	6	5	5	6	7
	0	2	2	4	—	—	—	6	5	2	6	2	3	4	2	3
3	3	0	3	2	—	—	—	5	3	17	1	4	4	3	4	7
	2	1	3	1	5	1	—	10	13	14	10	7	11	11	5	6
	1	2	5	3	4	4	3	19	17	1	12	8	5	10	9	7
	0	3	3	—	1	1	1	6	5	1	1	5	4	—	6	4
4	4	0	1	—	3	—	—	4	—	13	1	2	1	—	2	2
	3	1	2	1	3	—	1	7	11	14	11	7	5	8	9	8
	2	2	9	2	3	1	1	16	11	7	9	11	11	13	11	10
	1	3	5	4	1	—	1	11	12	1	8	6	7	6	4	7
	0	4	1	—	1	—	2	4	1	1	—	3	5	2	3	2
5	5	0	1	—	—	—	—	1	3	9	1	—	—	1	—	—
	4	1	4	2	5	1	—	12	2	12	8	5	5	5	6	6
	3	2	8	1	9	2	2	22	21	7	15	15	18	14	14	15
	2	3	6	5	3	1	3	18	9	8	11	17	10	12	16	11
	1	4	2	3	3	—	—	8	8	—	5	6	8	9	6	10
	0	5	3	—	2	—	—	5	—	—	3	—	2	2	1	1
6	6	0	—	—	—	—	—	—	1	6	—	—	—	1	1	—
	5	1	4	—	6	2	—	12	5	18	5	2	5	3	1	5
	4	2	3	3	6	2	1	15	8	7	10	5	8	8	6	10
	3	3	9	1	9	3	2	24	18	7	8	11	14	9	13	12
	2	4	7	3	6	3	2	21	6	—	7	13	4	9	7	5
	1	5	2	1	2	3	3	11	3	—	3	2	2	1	3	1
	0	6	—	2	—	—	—	2	2	—	—	—	—	2	2	—
7	7	0	—	—	—	—	—	—	—	8	—	—	—	—	2	1
	6	1	1	1	3	1	2	8	2	10	2	1	2	2	—	2
	5	2	5	1	1	—	1	8	9	9	5	5	5	8	4	7
	4	3	6	—	8	2	1	17	14	9	8	7	9	6	9	8
	3	4	5	2	7	3	5	22	8	3	11	8	7	7	6	7
	2	5	2	2	—	4	1	9	8	1	2	8	6	5	6	4
	1	6	1	—	—	—	—	1	3	—	1	1	1	2	3	1
	0	7	—	—	—	—	—	—	—	—	1	—	—	—	—	—

TABLE II (*continued*)

Litter size	D	r	1						2	3	4					
			I	II	III	IV	V	Total			Sex	a	wv	s	b	d
8	8	0	—	—	—	1	—	1	—	3	1	—	—	—	—	—
	7	1	1	—	2	—	1	4	—	3	1	1	2	1	1	1
	6	2	2	—	2	—	4	8	2	7	5	3	5	3	8	1
	5	3	4	5	10	4	4	27	3	4	8	4	6	9	5	12
	4	4	6	—	9	1	3	19	5	—	7	9	12	10	5	9
	3	5	6	2	6	2	6	22	4	—	5	12	4	6	9	6
	2	6	—	1	3	2	2	8	3	—	5	3	2	3	3	3
	1	7	2	—	1	3	—	6	—	—	—	—	1	—	1	—
	0	8	—	—	—	—	1	1	—	—	—	—	—	—	—	—
9	9	0	—	—	1	—	—	1	1	—	—	—	—	—	—	—
	8	1	—	—	—	1	—	1	—	—	—	—	—	—	—	—
	7	2	2	1	2	1	2	8	—	1	—	1	—	—	2	1
	6	3	5	1	3	5	2	16	3	3	4	2	6	2	7	5
	5	4	3	1	3	4	2	13	4	—	6	4	6	2	2	5
	4	5	2	1	4	2	3	12	2	—	4	4	2	7	3	3
	3	6	2	—	4	5	4	15	1	1	1	3	1	3	—	1
	2	7	4	1	1	2	1	9	1	—	—	1	—	1	1	—
	1	8	1	—	—	—	—	1	1	—	—	—	—	—	—	—
	0	9	—	—	—	—	—	—	—	—	—	—	—	—	—	—
10	9	1	—	—	—	—	—	—	—	1	—	—	—	—	—	—
	8	2	—	—	—	3	1	4	—	1	—	—	—	—	1	—
	7	3	6	—	2	2	—	10	—	—	1	—	—	—	—	1
	6	4	5	—	2	1	3	11	—	—	—	—	1	1	3	1
	5	5	5	1	—	1	—	7	—	1	1	1	1	1	—	1
	4	6	1	1	—	2	1	5	—	—	1	1	2	1	—	—
	3	7	2	1	2	2	—	7	—	—	1	1	—	1	—	—
	2	8	1	—	—	—	1	2	—	—	—	1	—	—	—	—
	1	9	—	—	—	1	1	2	—	—	—	—	—	—	—	—
11	9	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	8	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	7	4	—	—	—	—	—	—	—	—	—	—	—	1	—	—
	6	5	1	—	—	—	—	1	1	—	—	1	—	—	—	1
	5	6	2	—	—	—	—	2	—	—	—	—	1	—	—	—
	4	7	1	—	—	1	—	2	—	—	1	—	—	—	1	—
	3	8	1	—	—	—	—	1	—	—	—	—	—	—	—	—
	2	9	1	—	—	—	—	1	—	—	—	—	—	—	—	—
12	9	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	8	4	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	7	5	—	—	—	1	—	1	—	—	—	—	—	—	—	—
	6	6	—	—	—	1	1	2	—	—	—	—	—	—	—	—
	5	7	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	4	8	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	3	9	—	—	—	—	1	1	—	—	—	—	—	—	—	—
14	6	8	—	—	—	—	1	1	—	—	—	—	—	—	—	—

TABLE III

Exp.	Dominants	Recessives	Expectation	χ^2
1, I	567	577	572	0·087
1, II	162	198	180	3·600
1, III	528	457	492·5	5·118
1, IV	306	326	316	0·645
1, V	277	309	293	1·747
1, Total	1840	1867	1853·5	0·197*
2	685	681	683	0·012
3	899	299	299·5	0·001
4, sex	662	617	639·5	1·583
4, agouti	608	671	639·5	3·103
4, brown	648	631	639·5	0·226
4, spotting	636	643	639·5	0·038
4, wavy	658	621	639·5	1·070
4, blue	673	596	634·5	4·636

* This value of χ^2 is not the total of the five values given above it, but the value, having 1 degree of freedom, calculated from the totalled dominants and recessives of Exp. 1.

In all the cases the expectations are, of course, so large that we can use the classical χ^2 with complete confidence.

The calculation of χ^2 for each experiment from the data of Table II is rapid and simple. If a and b are the numbers of dominants and recessives, then in the case of a back-cross, where equality is expected, we multiply the numbers of litters containing a dominants and b recessives by $(a-b)^2$, sum the products for each value of s, and divide the sum by s. This gives the contribution to χ^2 made by litters of that particular size.

For example in the case of the total of Exp. 1 and litters of 6 mice the calculations are as follows:

Dominants	Recessives	$(a-b)^2$	n	$s\chi^2$
6	0	36	0	0
5	1	16	12	192
4	2	4	15	60
3	3	0	24	0
2	4	4	21	84
1	5	16	11	176
0	6	36	2	72
			85	584

Hence for litters of $s = 6$, $\chi^2 = \frac{584}{6} = 97{\cdot}\dot{3}$, the expected value as a result of random sampling being 85.

The results of applying this method to the data of Exp. 1 are given in Table IV, and compared with expectations in Table V. It will be seen that in every case χ^2

TABLE IV

	1, I		1, II		1, III		1, IV		1, V		1, Total	
♀	n	χ^2	n	χ^2	n	χ^2	n	χ^2	n	χ^2	n	χ^2
1	1	1·000	6	6·000	0	—	3	3·000	1	1·000	11	11·000
2	7	4·000	6	8·000	4	0·000	0	—	4	2·000	21	14 000
3	14	20·667	6	7·333	10	6·000	6	4·667	4	4·000	40	42·667
4	18	15·000	7	5·000	11	20 000	1	0 000	5	10·000	42	50·000
5	24	33·600	11	10 200	22	26·800	4	2·400	5	1·000	66	74·000
6	25	22 667	10	18 667	29	29·333	13	16·667	8	10·000	85	97·333
7	20	17·714	6	7·714	19	14·143	10	9·429	10	10·571	65	59·571
8	21	22·500	8	5·500	33	31·500	13	28·500	21	29·500	96	117·500
9	19	29·667	5	6·778	18	25·111	20	24·444	14	14·889	76	100·889
10	20	18·800	3	2·000	6	7 200	12	24·800	7	15·200	48	68·000
11	6	7·818	0	—	0	—	1	0·818	0	—	7	8 636
12	0	—	0	—	0	—	2	0·333	2	3 000	4	3·333
13	0	—	0	—	0	—	0	—	0	—	0	—
14	0	—	0	—	0	—	0	—	1	0·286	1	0·286
	175	193 433	68	77·192	152	160·087	85	115·058	82	101·446	562	647·215

TABLE V

Exp.	χ^2	n	χ^2-n	σ	d/σ
1, I	193·43	175	+18·43	16·87	+1·09
1, II	77·19	68	+ 9·19	9·96	+0·92
1, III	160·09	152	+ 8·09	15·83	+0·51
1, IV	115·06	85	+30·06	10·96	+2 74
1, V	101·45	82	+19·45	11·62	+1·67
1, Total	647·22	562	+85·22	30·12	+2·83

exceeds its expectation, that the excess is significant in the total and in 1, IV, and is very probably so in 1, V.

The effect of the corrections to Pearson's χ^2 is of interest. The variance of the total is reduced from $2n$, or 1124, to $2(n-\Sigma s_r^{-1})$ or 907·46. Thus σ is reduced from 33·52 to 30·12, and $\frac{\chi^2-n}{\sigma}$ is increased from 2·54 to 2·83. The normality of the distribution of χ^2 is also improved. β_1 is reduced from 0·0258 to 0·0079, and γ_2 from 0·0387 to 0·0097. The closeness of the approach to normality may be realized as follows. The variance of $\gamma_1 = \beta_1^{\frac{1}{2}}$ in a sample of N from a normal population is approximately $6/N$. Hence a value of $\beta_1 = 0{\cdot}0079$ would be found about once in three times in a sample of $6/\beta_1$ or 760 individuals from a normal population. It is in fact negligible.

In Exp. 1, II, which gave only 68 litters, $\beta_1 = 0{\cdot}0602$, $\beta_2 = 3{\cdot}00145$. The skewness is hardly worth considering in a test of significance.

Exps. 2 and 3 give the results shown in Table VI. To calculate the values of χ^2 in Exp. 3 we note that for a litter of s containing a dominants and b recessives

$$\chi^2 = \frac{(\frac{3}{4}s - a)^2}{\frac{3}{4}s} + \frac{(\frac{1}{4}s - b)^2}{\frac{1}{4}s} = \frac{(a-3b)^2}{3s}.$$

TABLE VI

	Exp. 2		Exp. 3	
s	n	χ^2	n	χ^2
1	11	11·000	14	12·667
2	28	24·000	21	24·667
3	38	34·000	33	30 333
4	35	27 000	36	44 000
5	43	39·000	36	46·133
6	43	48·667	38	31·556
7	44	42·857	40	53·714
8	17	13·500	17	12·667
9	13	21·889	5	9·370
10	0	0·000	3	4·667
11	1	0·091	0	0·000
Total	273	262·004	243	269·774

Hence if we multiply the number of each litter type by $(a-3b)^2$ and divide the total for each litter size by $3s$ we obtain the contribution of that litter size to χ^2.

The deviation in Exp. 2 is $-10{\cdot}996$, its standard error being 20·04. The deviation in Exp. 3 is $+26{\cdot}77$, its standard error being 18·89. Thus neither deviation is significant. We shall later have to consider the effect on χ^2 of selecting our material.

The results of applying the χ^2 test to Exp. 4 are given in Table VII. It will be seen that five out of the six values of χ^2 are less than their expectation. The variance is 353·41 except in the case of blue dilution, where it is 351·61. It will be seen that none of the deviations, taken by itself, is significant. The total value of χ^2 is 1277·38, its expectation being $1355 \pm 46{\cdot}03$. The deviation is $-1{\cdot}69$ times the standard error, which again is not significant. A considerably larger negative deviation would have suggested that the authors had suppressed a few aberrant families. An application of χ^2 to certain published work would, we are inclined to believe, give ground for such a suggestion.

We must next ask whether the large positive deviations of Exp. 1 can be explained. In order to analyse this experiment further the mice were grouped,

TABLE VII

s	n	χ^2					
		Sex	Non-agouti	Wavy	Spotting	Brown	Blue
1	4	4·000	4·000	4·000	4·000	4·000	4·000
2	11	14·000	10·000	12·000	12·000	10·000	8·000
3	24	13·333	32·000	10·667	16·000	34·667	37·333
4	29	23 000	33·000	36·000	22·000	33·000	31·000
5	43	48 600	26·200	39·000	45·400	32 600	39 000
6	33	32·667	22·667	26·667	40·000	37·333	26·000
7	30	29·429	26·000	27·143	32·857	39·714	34·000
8	32	39·000	24·500	32·500	24 000	38·000	20 000
9	15	6·111	11·444	7·889	8 778	15·889	9 667
10	4*	3·600	5·600	1·200	2 400	4·800	2·000
11	1	0·818	0·091	0·091	0 818	0·818	0·091
Total	226	214·558	195·502	197·157	208·253	250·821	211·091
d		−11·44	−30·50	−28·84	−17 75	+24·82	−13·91
d/σ		− 0·61	− 1·62	− 1·53	− 0 94	+ 1·32	− 0·74

* One family of 10 was not scored for blue dilution.

not in litters, but in families. The result is shown in Table VIII. The numbers of families, except in Exp. 1, I, and the total, are so small that the distribution of χ^2 is far from normal. And some of them are so small that the classical distribution is also inapplicable. It is clear, however, that the total χ^2 exceeds

TABLE VIII

Exp. 1, *litters grouped in families*

Exp.	Number of families	Number of mice	χ^2	$\chi^2 - n$	σ
1, I	107	1144	126 58	+19·58	13·56
1, II	26	360	33 03	+ 7·03	6·70
1, III	64	985	90·84	+26·84	10·72
1, IV	31	632	46·22	+15·22	7·54
1, V	25	586	33·74	+ 8·74	6·79
1, Total	253	3707	330·40	+77·40	21·13

its expectation by 3·66 times its standard error, and the divergences in Exps. 1, III and 1, IV must probably be regarded as significant. In each case, however, the divergence was mainly due to a single family. A family of 9 dominants and no recessive contributed 9 to the χ^2 of 1, III, and a family of 3 dominants and

14 recessives contributed 7·12 to the χ^2 of 1, IV. Were these families omitted no single experiment would yield a significant result, though their total would do so.

It is notable that when the litters are grouped by families, although n is reduced to 45 % of its former value, the excess of χ^2 is only reduced from 85·2 to 77·4. Thus there is little heterogeneity due to divergence of litters within families. We must look for a cause which affects families rather than litters. The following are possible causes:

(1) The presence of recessive lethal or sublethal genes linked with those segregating.

(2) The presence of monozygotic twins or multiplets.

(3) The existence of environmental factors which on some occasions favoured coloured mice, on others dilute mice, although on the whole they favoured neither.

(4) Abnormalities of meiosis leading to production of gametes in unequal numbers where equality was expected.

Only the first, and possibly the fourth of these, would affect families rather than litters. This is plausible on general grounds. The presence of the same lethal or sublethal gene in both parents is only likely as the result of inbreeding. There was inbreeding in all five parts of Exp. 1, but a perusal of Grüneberg's (1936) paper makes it clear that it was least in 1, I and 1, II, greater in 1, III, and greatest in 1, IV, and 1, V. Our results do not prove the segregation of lethal genes, but they are consistent with it. On the other hand such genes must have been rare or absent in Exps. 2, 3 and 4, although 2 and 3 involved a good deal of inbreeding and 4 a certain amount. It is hoped, by the application of this method to man, to obtain at least an upper limit to the frequency of lethals in human chromosomes, and possibly to obtain evidence suggesting their presence.

We must next consider the effect of selection on Exps. 2 and 3. In Exp. 2 the fully coloured parents were not always known to be heterozygous, and families of less than 7 were excluded unless the parent was known to be so. We may, however, have neglected some families of 7 or over derived from a heterozygous parent because they contained no dilute animals. The probability of such a family is 2^{-7} or less. Thirty-six families of 7 or more were inferred to be derived from a heterozygous parent because they contained one or more extreme dilute mice. The probability that such a group taken at random from the progeny of a heterozygote and a homozygote would have contained a family with no extreme dilute mice is 0·0365. A family of 7 would have contributed 7 to χ^2. Thus the value of χ^2 should be increased by about 0·25 of a unit to compensate for the effect of selection.

Similarly in Exp. 3 there were 30 families ranging from 16 to 35 in number which were inferred to be derived from matings of two heterozygotes because they contained at least one extreme dilute mouse. The probability that a family of n derived from two heterozygotes will include no recessive is $(\frac{3}{4})^n$. This is equal

to 0·0100 or less. Among 30 families of the given sizes the probability that one has been excluded on this ground is 0·0838. Such a family would contribute fairly heavily to χ^2. For example a family of 18 would contribute 6. It is concluded that the selection practised has probably reduced the value of χ^2 by less than half a unit. The data have therefore been legitimately used.

SUMMARY

The χ^2 test has been applied to data on Mendelian segregation on 1304 mouse litters containing 7550 mice. As some litters were scored for as many as six characters we have effectively 2433 litters, and 13935 mice. In 7 experiments χ^2 exceeded its expectation, in 6 it fell below it. None of the negative deviations was significant, but one of the positive ones was so. This is tentatively ascribed to the effect of recessive lethal genes in upsetting Mendelian segregation in an inbred population.

REFERENCES

DE WINTON, D. & HALDANE, J. B. S. (1933). "The genetics of *Primula sinensis*. II." *J. Genet.* XXVII, 1–44.

FISHER, R. A. & MATHER, K. (1936). "A linkage test with mice." *Ann. Eugen.*, Camb., VII, 265–80.

GRÜNEBERG, H. (1936). "Further linkage data on the albino chromosome of the house mouse." *J. Genet.* XXXIII, 255–65.

HALDANE, J. B. S. (1937). "The exact value of the moments of the distribution of χ^2, used as a test of goodness of fit, when expectations are small." *Biometrika*, XXIX, 133–43.

MISCELLANEA

(i) An Application of the Method of Maximum Likelihood

By WALTER A. HENDRICKS

Bureau of Animal Industry, United States Department of Agriculture

In a recent paper Pearson (1936) presented some rather critical observations on the interpretation of results derived from applications of the method of maximum likelihood. The author of the present paper has no desire to attempt a theoretical justification of the method of maximum likelihood. He can, however, add another scrap of evidence to the mounting total which has done much to justify the method by the principle of induction.

Suppose that a set of 17 individuals, drawn from some universe, may be divided into four classes. Let the numbers of individuals in the four classes be 3, 7, 2, and 5, respectively. Assume that a hypothesis regarding the universe causes us to expect twice as many individuals in the second class as in the first, and twice as many in the fourth class as in the third. In other words, assume that the respective probabilities for the four classes are p_1, $2p_1$, p_2, and $2p_2$. Let it be required to obtain estimates of p_1 and p_2 from the observed distribution of 17 individuals.

The investigator equipped with a knowledge of the more elementary aspects of simple sampling would doubtless proceed by reasoning that the probability of the occurrence of an individual in either the first or second class is equal to $3p_1$. He would equate this to the observed proportion, 10/17, of individuals in the two classes and solve the resulting equation, thus obtaining 10/51 or ·19607843 for his estimate of p_1. He would then apply the same process to the data in the third and fourth classes, or he would make use of the relation,

$$3p_1 + 3p_2 = 1, \qquad \text{......(1)}$$

to reach the conclusion that p_2 is equal to 7/51 or ·13725490.

It is of interest to note that this solution, which would appeal to the experienced investigator, is exactly that to which the method of maximum likelihood leads.

In applying the method of maximum likelihood to the above problem, we are required to determine p_1 and p_2 so as to give the maximum value to the quantity, L, defined by

$$L = 3 \log p_1 + 7 \log 2p_1 + 2 \log p_2 + 5 \log 2p_2, \qquad \text{......(2)}$$

subject to the condition imposed by equation (1). This is equivalent to determining p_1, p_2, and λ_1 in such a manner as to give the maximum value to the quantity, Y, defined by

$$Y = 10 \log p_1 + 7 \log p_2 + \lambda_1 (3p_1 + 3p_2 - 1). \qquad \text{......(3)}$$

The required values of p_1, p_2, and λ_1 are given by the solution of the equations:

$$\left.\begin{aligned} 10/p_1 + 3\lambda_1 &= 0, \\ 7/p_2 + 3\lambda_1 &= 0, \\ 3p_1 + 3p_2 &= 1, \end{aligned}\right\} \qquad \text{......(4)}$$

from which the values of p_1 and p_2 are found to be 10/51 and 7/51, respectively, exactly as before.

It is also of some interest to obtain estimates of p_1 and p_2 in such a manner that the familiar criterion of goodness of fit, χ^2, as applied to the comparison of observed and theoretical frequencies in the four classes, shall have its minimum value.

The value of χ^2 is given by the relation

$$\chi^2 = 9/17p_1 + 49/34p_1 + 4/17p_2 + 25/34p_2 - 17. \qquad \text{......(5)}$$

To render this quantity a minimum, subject to the condition imposed by equation (1), we may determine p_1, p_2, and λ_2 in such a manner as to minimize the quantity, z, defined by

$$z = 67/p_1 + 33/p_2 + \lambda_2(3p_1 + 3p_2 - 1). \qquad \text{......(6)}$$

The required values of p_1, p_2, and λ_2 are given by the solution of the equations,

$$\left.\begin{aligned} -67/p_1^2 + 3\lambda_2 &= 0, \\ -33/p_2^2 + 3\lambda_2 &= 0, \\ 3p_1 + 3p_2 &= 1, \end{aligned}\right\} \qquad \text{......(7)}$$

from which the values of p_1 and p_2 are found to be ·19586990 and ·13746343, respectively.

These estimates of p_1 and p_2 differ very little from those obtained by the two preceding methods. However, they are different. This is not surprising, since the application of the χ^2 test to data such as those under consideration involves certain well-known approximations. The fact that the differences are rather small is in agreement with results obtained by Fisher (1934) in a similar comparison of methods of estimation.

The most interesting feature of the results presented in this paper is the fact that the method of maximum likelihood, as applied to the present problem, led to results which are in exact agreement with those obtained as natural consequences of the established theory of simple sampling.

REFERENCES

Fisher, R. A. (1934). *Statistical Methods for Research Workers*, 5th ed. Edinburgh and London: Oliver and Boyd.

Pearson, Karl (1936). "Method of moments and method of maximum likelihood." *Biometrika*, xxviii, 34–59.

(ii) Maximum Likelihood and Methods of Estimation

By E. S. Pearson

In the preceding Note Mr Walter A. Hendricks has given an example of the difference between the result of applying the methods of maximum likelihood and of minimum χ^2. The illustration is interesting and suggestive although I do not know whether it can be described as making any contribution towards *justifying* the general application of the method of maximum likelihood. When two or more alternative methods of estimation are available, we may choose between them in many ways, e.g. by an appeal to intuition or by an appeal to practical expediency. On both these counts the method of maximum likelihood would, in the present instance, score points over the method of minimum χ^2. But when we turn to the alternative methods of fitting frequency curves, with which my father was concerned in the paper referred to, e.g. (i) by maximum likelihood; (ii) by minimum χ^2; (iii) by moments, the choice is far less easy to make. From the point of view of any commonly experienced sense of intuition, I fancy there can be no unique answer; from the point of view of practical expediency, in many cases the moment method clearly wins.

As I have mentioned elsewhere in this *Journal*,* it is I think the practical worker who will give the final casting vote between alternative theoretical methods, basing his decision on considerations of practical utility. In the growing complexity of mathematical statistics it is, however, often difficult for him to make his choice without the aid of simple guiding principles which appeal to his intuition. The concept of maximum likelihood involves such a principle. To assign to an unknown probability, p, that value which, if it were correct, would make the occurrence of the observed result more likely than any other value, is a procedure which has a simple intuitional appeal. But clearly there may be other guiding principles which are equally useful to follow.

* Pp. 53–64 above.

Consider Mr Hendrick's illustration. It is almost certain that neither of the 8 decimal-place estimates of p_1 which he gives, $\cdot 1960{,}7843$ and $\cdot 1958{,}6990$, is the true population value. What would be of more use to the practical man than either of these would be a rule for obtaining from the sample an upper and lower estimate for p_1, say $\bar{p}_1$ and $\underline{p}_1$, such that the statement

$$\underline{p}_1 \leqslant p_1 \leqslant \bar{p}_1 \qquad \ldots\ldots(1)$$

might be made with a certain measure of confidence, expressed by the risk of error, say α, involved in the statement. In so far as the method permits the adjustment of the values of α and of the breadth of the interval $(\underline{p}_1, \bar{p})$, it would have a greater appeal than any method which only provides a single-valued estimate of p_1. If now two alternative methods are available, both enabling us to estimate an interval $(\underline{p}_1, \bar{p})$, we may make a choice between them, basing that choice upon a comparison of (*a*) the risk of error involved in the statement (1), and (*b*) the breadth of the intervals.

In this twofold principle we have gone beyond that involved in the simple choice of p_1 by maximizing the likelihood. That method may form a part of the process of determining the interval $(\underline{p}_1, \bar{p}_1)$, but it is now the means to an end, not the end itself.

When we come to the more complex problem of curve fitting, it is far less clear what result will be most useful to the practical worker. Starting from a given functional equation

$$y = p(x \mid \theta_1, \theta_2, \ldots, \theta_c) \qquad \ldots\ldots(2)$$

involving c unknown parameters θ, it is clear that a possible principle of estimation is to assign to the parameters the values, which if they were the population values, would make the occurrence of the observed sample more likely than any other set of parametric values But many have felt that there is something remote about the abstract conception involved. If the maximum likelihood procedure leads to estimates T_i of θ_i $(i=1, 2, \ldots, c)$ which in random sampling have smaller standard errors than any other form of estimates T_i, a result has been reached with a more direct practical appeal. The achievement would be greater still if a method were available for determining upper and lower limits $\bar{T}_i$ and $\underline{T}_i$ so that the statement

$$\underline{T}_i \leqslant \theta_i \leqslant \bar{T}_i (i=1, 2, \ldots, c) \qquad \ldots\ldots(3)$$

could be made jointly with regard to the c parameters, with a given risk of error. But in the present state of development of the theory of maximum likelihood, as applied to the fitting of frequency curves to samples of *finite* size, can it be said that such results have been achieved?

It must also be remembered that in so far as the statistician wishes to use his frequency curve for graduation purposes, the agreement between observation and fitted curve throughout the range of significant frequency may make a more direct and simpler appeal to him than any information regarding the values of the parameters, θ_i. The quantity

$$\chi^2 = \sum_s \left(\frac{n_s - m_s}{\sqrt{m_s}}\right)^2, \qquad \ldots\ldots(4)$$

summing up the relative discrepancy, may be more closely correlated with his conception of goodness of fit than any measure based on the likelihood, L, or on the reliability of the estimated parameters

Finally, the question of practical utility must play a dominating part; at present the method of fitting by moments, making where necessary certain empirical adjustments, does provide a practical working tool. Until far more exploratory work has been carried out on the application of the method of maximum likelihood in fitting frequency curves, it is quite impossible to attempt any assessment of its final value.

It was some of these considerations, expressed perhaps in a different form, that I believe my father had in mind when challenging the claim that the method of maximum likelihood is the only efficient method of fitting frequency curves.

(iii) A Note on Unbiased Limits for the Correlation Coefficient

By F. N. DAVID

We shall assume that a sample of size n has been randomly drawn from a normal bivariate population such as

$$p(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left\{\frac{(x-\xi_1)^2}{\sigma_1^2} + \frac{2\rho(x-\xi_1)(y-\xi_2)}{\sigma_1\sigma_2} + \frac{(y-\xi_2)^2}{\sigma_2^2}\right\}},$$

and that we are interested in ρ, the coefficient of correlation between x and y in the population. It is possible that we may require answers to two questions:

(1) Given the sample of size n, are these observations consistent with the hypothesis that $\rho = \rho_0$, where ρ_0 is some specified value?

(2) Given the sample of size n, how may we calculate ρ_1 and ρ_2, so that, subject to the risk which we are willing to undertake, the interval ρ_1 to ρ_2 will cover the true population value as often as possible?

There are other questions which we may ask, and these will be dealt with fully in the introduction to "Tables of the Ordinates and Probability Integral of the distribution of r" which

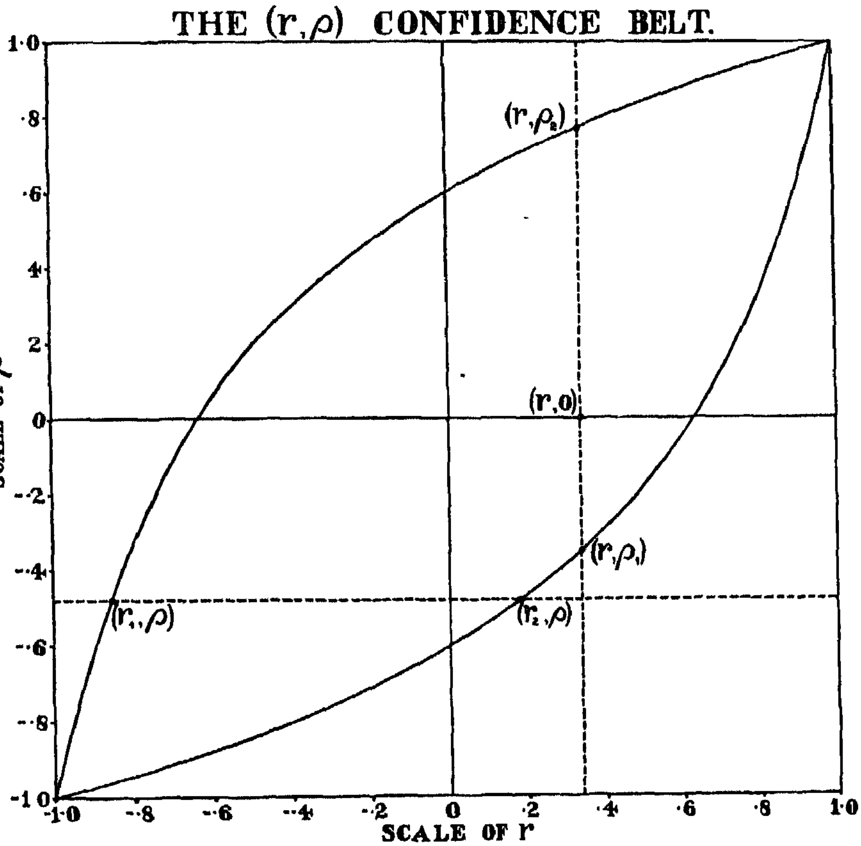


Fig. 1.

is shortly to be published. In the present note we shall confine ourselves solely to those questions in which a consideration of "bias" is important.

The methods of answering these questions may be illustrated with the help of the diagram on p. 157. If for a given n for each value of ρ in the range -1 to $+1$ we calculate from the probability distribution of r, say $p_n(r \mid \rho)$, limits r_1 and r_2 such that

$$\int_{-1}^{r_1} p_n(r \mid \rho)\, dr = \alpha_1, \quad \int_{r_2}^{+1} p_n(r \mid \rho)\, dr = \alpha_2, \qquad \ldots\ldots(1)$$

where $$\alpha_1 = \alpha_2 \qquad \ldots\ldots(2)$$

and $$\alpha_1 + \alpha_2 = \alpha = \text{constant}, \qquad \ldots\ldots(3)$$

then the points (r_1, ρ) and (r_2, ρ) will fall on two curves enclosing a lozenge-shaped belt as shown.* Accordingly when dealing with question (1), if we decide to reject the hypothesis tested, i.e. that $\rho = \rho_0$, whenever the point (r, ρ_0) falls outside the belt, we shall run a risk equal to α of rejecting the hypothesis when it is true Question (2) may be answered with the aid of the same diagram. The point $(r, \rho = 0)$ is plotted and a line parallel to the axis of ρ is drawn through it. Suppose that this line cuts the belt in the points (r, ρ_1) and (r, ρ_2). Then we know that the interval ρ_1 to ρ_2 will cover the true value of ρ in the population in $100(1-\alpha)$ percentage of cases (Neyman, 1934). It will be noticed that the risk of error, α, would remain the same if equation (3) held, but not equation (2). Thus it is possible to obtain an infinite variety of belts satisfying (3) by following different principles in the determination of r_1 and r_2, or α_1 and α_2.

Neyman & Pearson (1936), when discussing questions similar to question (1), showed that in certain skew distributions the limits obtained by taking equal tail areas led to a curious anomaly. In such cases they found that the hypothesis tested was more likely to be rejected when it was true, than when in fact an alternative hypothesis was true. A test leading to such consequences Neyman & Pearson termed "biased". It is the object of the present note to find unbiased limits for r, and to compare them with the limits found by taking equal tail areas.

The probability distribution of r, for any n and ρ may be written as follows:

$$p_n(r \mid \rho) = \frac{(1-\rho^2)^{\frac{n-1}{2}}}{(n-3)\,!} \frac{(1-r^2)^{\frac{n-4}{2}}}{\pi} \frac{d^{n-2}}{d(r\rho)^{n-2}} \left(\frac{\text{arc cos}\,(-\rho r)}{\sqrt{1-\rho^2 r^2}}\right). \qquad \ldots\ldots(4)$$

Following the procedure of Neyman & Pearson we see that an unbiased test will be obtained by solving equations (5) and (6) for r_1 and r_2.

$$\int_{r_1}^{r_2} p_n(r \mid \rho)\, dr = 1-\alpha, \qquad \ldots\ldots(5)$$

$$\frac{d}{d\rho} \int_{r_1}^{r_2} p_n(r \mid \rho)\, dr = 0, \qquad \ldots\ldots(6)$$

where α is chosen, as is customary, according to the risk we are willing to undertake of rejecting the hypothesis as false when it is true. Differentiating (6) with respect to ρ we get

$$0 = \frac{-\rho(n-1)}{1-\rho^2} \int_{r_1}^{r_2} \frac{(1-\rho^2)^{\frac{n-1}{2}}}{(n-3)\,!} \frac{(1-r^2)^{\frac{n-4}{2}}}{\pi} \frac{d^{n-2}}{d(r\rho)^{n-2}} \left(\frac{\text{arc cos}\,(-\rho r)}{\sqrt{1-\rho^2 r^2}}\right) dr$$
$$+ \int_{r_1}^{r_2} \frac{(1-\rho^2)^{\frac{n-1}{2}}}{(n-3)\,!} \frac{(1-r^2)^{\frac{n-4}{2}}}{\pi} r \,.\, \frac{d^{n-1}}{d(r\rho)^{n-1}} \left(\frac{\text{arc cos}\,(-\rho r)}{\sqrt{1-\rho^2 r^2}}\right) dr,$$

that is

$$0 = \frac{-\rho(n-1)(1-\alpha)}{(1-\rho^2)} + \int_{r_1}^{r_2} \frac{(1-\rho^2)^{\frac{n-1}{2}}}{(n-3)\,!} \frac{(1-r^2)^{\frac{n-4}{2}}}{\pi} r \,.\, \frac{d^{n-1}}{d(r\rho)^{n-1}} \left(\frac{\text{arc cos}\,(-\rho r)}{\sqrt{1-\rho^2 r^2}}\right) dr. \qquad \ldots\ldots(7)$$

* The diagram illustrates approximately the case $n = 10$, $\alpha = 0{\cdot}05$; charts of these "confidence belts" for varying n and α will be given in the publication referred to above.

Integrate (7) by parts, remembering that

$$\frac{d}{dr}\left[\frac{d^{n-1}}{d(r\rho)^{n-1}}\left(\frac{\arccos(-\rho r)}{\sqrt{1-\rho^2r^2}}\right)\right] = \rho\frac{d^n}{d(r\rho)^n}\left(\frac{\arccos(-\rho r)}{\sqrt{1-\rho^2r^2}}\right)$$

and we get

$$\int_{r_1}^{r_2}(1-r^2)^{\frac{n-4}{2}}r\,.\,\frac{d^{n-1}}{d(r\rho)^{n-1}}\left(\frac{\arccos(-\rho r)}{\sqrt{1-\rho^2r^2}}\right)dr = -\left[\frac{(1-r^2)^{\frac{n-2}{2}}}{n-2}\right.$$

$$\left.\times\frac{d^{n-1}}{d(r\rho)^{n-1}}\left(\frac{\arccos(-\rho r)}{\sqrt{1-\rho^2r^2}}\right)\right]_{r_1}^{r_2}+\frac{\rho}{n-2}\int_{r_1}^{r_2}(1-r^2)^{\frac{n-2}{2}}\frac{d^n}{d(r\rho)^n}\left(\frac{\arccos(-\rho r)}{\sqrt{1-\rho^2r^2}}\right)dr. \quad \ldots.(8)$$

From (4) we see that we may write

$$p_{n+1}(r\mid\rho) = \frac{(1-\rho^2)^{\frac{n}{2}}}{(n-2)!}\frac{(1-r^2)^{\frac{n-3}{2}}}{\pi}\frac{d^{n-1}}{d(r\rho)^{n-1}}\left(\frac{\arccos(-\rho r)}{\sqrt{1-\rho^2r^2}}\right), \quad \ldots\ldots(9)$$

$$p_{n+2}(r\mid\rho) = \frac{(1-\rho^2)^{\frac{n+1}{2}}}{(n-1)!}\frac{(1-r^2)^{\frac{n-2}{2}}}{\pi}\frac{d^n}{d(r\rho)^n}\left(\frac{\arccos(-\rho r)}{\sqrt{1-\rho^2r^2}}\right). \quad \ldots.(10)$$

Substituting (10), (9) and (8) in (7) we get

$$(1-\alpha) = \int_{r_1}^{r_2}p_{n+2}(r\mid\rho)\,dr - \frac{\sqrt{1-\rho^2}}{\rho(n-1)}\left[p_{n+1}(r\mid\rho).\sqrt{1-r^2}\right]_{r_1}^{r_2}, \quad \ldots.(11)$$

or making use of (5)

$$\int_{r_1}^{r_2}p_n(r\mid\rho)\,dr = \int_{r_1}^{r_2}p_{n+2}(r\mid\rho)\,dr - \frac{\sqrt{1-\rho^2}}{\rho(n-1)}\left[p_{n+1}(r\mid\rho).\sqrt{1-r^2}\right]_{r_1}^{r_2} \quad \ldots\ldots(12)$$

Hence solving for r_1 and r_2 from (5) and (12) we should get the r_1 and r_2 which are the unbiased limits for r for a given n and ρ.

An algebraical solution of (5) and (12) proved elusive. Accordingly it was decided to solve equations (5) and (12) for r_1 and r_2 by means of trial and error, given one specific size of sample. The size of sample chosen was $n = 10$; this because it is unlikely that the correlation coefficient would be worked out for a sample of less than 10 observations. For all samples of more than 10 observations the bias, if any, would be expected to be less than that for the sample of 10, since the distribution curves of r tend slowly to normality.

The method of procedure was as follows: $1-\alpha$ was chosen as 0·95, and the first value of ρ to be considered was $\rho = 0{\cdot}5$. Using the unpublished tables (David, 1937) of the probability integral of r, r_1 and r_2 were found by backward interpolation into the tables, for

$$\alpha_1 = 0{\cdot}025 = \alpha_2,\ n = 10,\ \rho = 0{\cdot}5,\ \alpha = 0{\cdot}05.$$

Equal tail areas give $r_1 = -0{\cdot}1556$; $r_2 = 0{\cdot}8673$. The right-hand side of equation (11) was evaluated using these values of r_1 and r_2. Instead of being equal to 0·95, as it would have been had there been no bias, it was equal to 0·9512. Several other values of α_1 and α_2 were tried. Finally taking $\alpha_1 = 0{\cdot}0245$ and $\alpha_2 = 0{\cdot}0255$, by backward interpolation equation (5) gave $r_1 = -0{\cdot}1591$; $r_2 = 0{\cdot}8664$. Evaluating the right-hand side of (11), using these values for r_1 and r_2 it was found to equal 0·95. Hence the r_1 and r_2, given by $\alpha_1 = 0{\cdot}0245$ and $\alpha_2 = 0{\cdot}0255$, are the unbiased limits for r, and it is seen that the bias is very small.

The distribution curve of r for $\rho = 0$ is symmetrical, so in this case the limits r_1 and r_2 obtained by taking equal tail areas will also be the unbiased limits. The distribution curves of r gradually become asymmetrical as ρ increases. It therefore seems reasonable to suppose that the bias gradually increases with the asymmetry. We have investigated the case where $\rho = 0{\cdot}5$. Let us now consider the bias when $\rho = 0{\cdot}8$. The same method was carried out as before and we obtained

$$n = 10,\quad \rho = 0{\cdot}8\begin{cases}\alpha_1 = 0{\cdot}025 & \alpha_2 = 0{\cdot}025 \quad r_1 = +0{\cdot}4003 \quad r_2 = 0{\cdot}9550^+\\ \alpha_1 = 0{\cdot}0242 & \alpha_2 = 0{\cdot}0258 \quad r_1 = +0{\cdot}3959 \quad r_2 = 0{\cdot}9545^+.\end{cases}$$

We see that there is a greater difference between the tail areas than for $\rho = 0.5$, but that for the r's the difference between the unbiased values and those obtained from equal tail area limits is only slightly increased for r_1 and slightly decreased for r_2. The area under the distribution curves for both $\rho = 0{\cdot}5$ and $\rho = 0{\cdot}8$ is unity, but the standard deviation for the curve $\rho = 0{\cdot}8$ is less than that for the curve $\rho = 0{\cdot}5$. Hence an alteration in our tail areas for the curve $\rho = 0{\cdot}8$, will mean much less change in the limits for r than for the same alteration in the tail areas of the curve for $\rho = 0{\cdot}5$. Our result accordingly seems reasonable, and we may therefore conclude that the unbiased limits for r follow very closely those limits which are found by taking equal tail areas.

In answering question (1) we see that in testing the hypothesis $\rho = \rho_0$, with admissible alternative hypotheses $-1 < \rho < \rho_0$ and $\rho_0 < \rho < +1$, if we used the limits for r obtained from equal tail areas we should reject the hypothesis tested as false according to the prescribed risk, but that there is a possibility that we should reject some other wrong hypothesis even less frequently.

In answering question (2) we may note that, between the points $(r_1, \rho = 0)$ and $(r_2, \rho = 0)$, the equal tail area interval for ρ is actually narrower than the unbiased interval, while for $(-1, \rho = 0)$ to $(r_1, \rho = 0)$, and $(r_2, \rho = 0)$ to $(+1, \rho = 0)$ the intervals practically coincide. It might therefore be asked why the unbiased interval should be chosen. The risk of the interval ρ_1 to ρ_2 failing to cover the true population value of ρ is fixed in both cases as $0{\cdot}05$, and since the unbiased interval is the greater over a very large range of r, why not choose the other? The answer is found in the definition of bias. In the case of the unbiased interval we see that this interval is chosen to cover the true population value 95 times in 100, and any other wrong value fewer times. In the case of the equal tail areas the interval is chosen to cover the true population value 95 times in 100, but it may cover some other wrong value more times than it does the true value. In the case of the r-distribution this discussion is, of course, theoretical, since the bias is so small, but it is conceivable that the point will prove important in other distributions.

It was expected that the bias in the distribution of r would prove to be small, because of Prof. Fisher's z' transformation (Fisher, 1921) for r. This transformation is nearly perfect, and transforms the asymmetrical curves of r into a series of normal curves. We should take equal tail areas from these normal curves in order to get unbiased limits for z', and therefore on transforming back we should expect to take equal tail areas from the r-curves. Since the transformation is not quite perfect we should expect a slight bias, which is what is actually found.

REFERENCES

NEYMAN, J. (1934). *J. R. statist. Soc.* XCVII, 589–93.

NEYMAN, J. & PEARSON, E. S. (1936). *Statist. Res. Mem.* I, 1–37.

DAVID, F. N. (1937) "Tables of the Ordinates and Probability Integral of the distribution of r in small samples." (Shortly to be published.)

FISHER, R. A. (1921). *Metron*, I, Part 4, 14.

Vol. XXIX. Parts III and IV February, 1938

BIOMETRIKA

A JOURNAL FOR THE STATISTICAL STUDY OF BIOLOGICAL PROBLEMS

FOUNDED BY
W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

EDITED BY
EGON S. PEARSON

IN CONSULTATION WITH
HARALD CRAMÉR J. B. S. HALDANE
MAJOR GREENWOOD G. M. MORANT
JOHN WISHART

ISSUED BY THE BIOMETRIKA OFFICE
UNIVERSITY COLLEGE, LONDON
AND PRINTED AT THE
UNIVERSITY PRESS, CAMBRIDGE

PRINTED IN GREAT BRITAIN

Reprinted by offset-litho 1952

Karl Pearson: 1910

VOLUME XXIX FEBRUARY, 1938 PARTS III AND IV

KARL PEARSON

An Appreciation of some Aspects of his Life and Work

By E S. PEARSON

Part II. 1906–1936*

1906–1911

The year 1906 was a dividing point in Pearson's life; that he felt it so himself there is much evidence to be found in his letters. The thirteen years since 1893 had been marked by a growing personal friendship and scientific collaboration between Weldon and himself which could never be replaced. Together they had wrestled with the development of a new mathematical technique and had shown with abundant illustration how necessary was its application in many of the fundamental problems of biology. They had met fierce and sometimes unscrupulous opposition and had faced it together, Weldon with his dashing cavalry charges into the foe, Pearson with his heavier artillery. Since the former had moved to Oxford, there had been a continual exchange of correspondence between them on the problems with which each was concerned at the moment; letters full of the excitement of some new discovery or of frank criticism of each other's ideas. Of Pearson's letters to Weldon only a short series written about 1900 seems to have been preserved, but these throw so much light on the place of that friendship in the history of biometry that I shall quote two of them now, even if they belong strictly to an earlier period†.

* The first part of this memoir appeared in *Biometrika*, XXVIII (1936), pp 193—257. The two parts will shortly be re-issued together in book form.

† This group of letters only came into my possession after the publication of the first half of this memoir.

7, Well Road,
Hampstead, N.W.
April 23, 1900.

My dear Weldon,

I got home to-day to find your letter. I am awfully sick at getting back into this loathsome Town and harness, when there is so much to be seen and observed in the country. A competent man tells me that the Yaffle is the green woodpecker and that the peasants never call anything but the lesser or spotted woodpecker, a woodpecker. Here it is intensely hot and nothing but house-painters in view down the backs of houses. At Gatwick I could watch from my work-table two snipe preparing to nest, two pairs of plovers ditto, all in a marshy bit of common at the bottom of our small garden. Also a wryneck in a box within an inch of my window. The owner encourages birds by sticking up pots and boxes for them, and the pools on the common abound in water birds I don't know even the names of. Now I must come and sit here until July when it will be too late to see anything. What brutes those Oxford Electors were to condemn me to endless years of London!*

I am afraid the poppies are giving you endless trouble. I shall not get across to Highgate to look at my own sowings till this week end. But I shall hear something from Oliver, Tansley and Macdonell of theirs. They are all far less elaborate though than yours! I will send you as soon as I can the averages of the selected poppies.

I have written to 11 persons for 10 house sparrow nests. Suppose 50 per cent come up to the scratch I shall get 50 nests of 5 eggs or more apiece. I don't think I can get by bothering everybody I know more than 70. Can you provide 30? Latter of Charterhouse who is going to try and do ten suggested that museums often have a number of clutches of eggs of each species. What are your longest series at Oxford? How many plovers have you? Plover give too few eggs but by taking more individual nests one might make it up.

Yours always sincerely,
K. Pearson.

7, Well Road.
May 3rd, 1901.

My dear Weldon,

I am too sleepy to write much, so here are a string of statements:

(i) I have been nigh to Gloucester today and found rooms on the Cotswolds near Bisley. I think by biking to Fairford it ought to be possible to get to Oxford. This is something off one's mind. I hasten to tell you in case you should be asking your friends about places for us, and wasting your precious time.

(ii) The correlation of barometric height between Bodø, the northernmost Norwegian Station, and Valentia is hardly sensible, but between Bodø and Funchal it is sensibly *negative*!

(iii) The value of least girth to the breadth has, on the average of 700 house sparrow eggs, the value of 3·15. Do you think this a reasonable approximation to the value of π?

* This is a reference to Pearson's unsuccessful application for the Savilian and Sedleian Chairs at Oxford in 1897 and 1899 respectively; see Part I of this memoir, p. 224.

(iv) If you will send me particulars as to what you want done re Italian statistics, I think I can set somebody on to it.

(v) We have found out that the interval between births and the difference between duration of life of elder and younger brothers or sisters are sensibly correlated. The elder lives as a rule 4 years longer. In other words *order of birth* influences longevity. I have suggested in this final paper that much of the variability one finds in an array of brethren will be found like longevity to be correlated with order of birth. Is not this akin to differences in early and late flowers?

(vi) Have you read Guarta (?) on crossing of white and waltzing mice? I have only seen Davenport's account of it, but it seems as if the percentages were based on too few cases to be of real value. Still it appears to resemble Mendel's case of pea hybrids.

(vii) It is not for mathematical formulae, but to give large correlation tables, that the large size of *Biometrika* is worth fighting for. I think we must be assured that the Press don't intend to take a big publisher profit off us!

Yours always sincerely,

K. P.

Please mention to the Press that we shall want articles in *foreign tongues*.

The inspiration and encouragement gained from personal contact, largely during those holiday meetings in the country, had been great too. "You have hardly realised and I don't think *he* did," Pearson wrote in April 1906 to Mrs Weldon, "how mentally refreshing it was to me being near him for a few weeks and how it sent me back fit for work with new vigour and new ideas. He always gave me courage and hope to go on...."

In the years to follow Pearson stood alone, the leader of a cause, who must plan its action and fight its battles. It is true that for five years more he gained much from an ever growing friendship with Francis Galton, but neither the counsel of an old man nor the help of a body of younger followers could replace the comradeship that was lost. Perhaps more than at any moment in his life, in this black year of 1906, he needed to draw on that fund of courage which he possessed.

While it is easy to place too much emphasis on a classification of a life's work into periods, it is I think justifiable to associate the period 1906–1914 with Pearson's foundation of a research institute where the ideas and methods, thrown up in rapid succession during the previous years of excitement and discovery and dealing with investigations as yet regarded almost as hobbies, could be developed under more secure conditions into an established branch of science. It seems well therefore to preface this section of the memoir with some account of the origin of the Biometric and Eugenics Laboratories, which were later, in 1911, to form Pearson's Department of Applied Statistics, and to make clear how he regarded their relationship and their purpose.

With the steadily increasing application in biology of quantitative methods of comparison it is likely that the term biometry may come to be used in a wider or different sense than that originally understood. It is therefore of some interest

to record what Pearson, writing in 1920, regarded as the function of his Biometric Laboratory *.

"The origin of the Biometric Laboratory must be sought in the year 1895, when the present Galton Professor gave his first course on the mathematical theory of statistics—probably the first course given on the modern mathematical theory at all—to two or three postgraduate students, one of whom is now Reader of Statistics in the University of Cambridge. From that year the statistical course became annual, and as the field of this form of investigation had been very little worked, a school sprung up which has since been recognised as the 'Biometric School,' and the group of workers, occupying a small room at University College, later termed the Biometric Laboratory, issued a long series of memoirs, which formed the basis of the English school of mathematical statistics. The object of this school was to make statistics a branch of applied mathematics with a technique and nomenclature of its own, to train statisticians as men of science, to extend, discard or justify the meagre processes of the older school of political and social statisticians, and in general to convert statistics in this country from being the playing field of *dilettanti* and controversialists into a serious branch of science, which no man could attempt to use effectively without adequate training, any more than he could attempt to use the differential calculus, being ignorant of mathematics. This task was a very arduous one, for statistics in one form or another are fundamental in nearly every branch of science in precisely the same manner as mathematics are fundamental in astronomy and physics. Inadequate and even erroneous processes in medicine, in anthropology, in craniometry, in psychology, in criminology, in biology, in sociology, had to be criticised, not for the pleasure of controversy, but with the aim of providing those sciences with a new and stronger technique. The battle has lasted for nearly twenty years, but there are many signs now that the old hostility is over and the new methods are being everywhere accepted."

It is clear from this statement that Pearson regarded his Biometric Laboratory as essentially a centre for training postgraduate workers in a new branch of exact science and for the application of the methods learnt, partly as illustrations of technique, in a variety of different directions. The field of application for which the technique had been devised was that of biology, and for this reason the word *βίος* had been associated with *μέτρον*, but Pearson had already illustrated the use of the new methods in meteorology and astronomy. Writing to Galton in 1908 ((**18**) III A, p. 333) he said:

"I have so much in hand that to close one phase of my work only means more progress in other phases. I should only feel sad if something were to happen which closed *all* phases of my work. Why, if Eugenics and even Biometry were closed down, I should turn to Astronomy with all my energy and time; I know how badly statistical knowledge is needed for problems therein!"

In his anxiety to leave Galton free to change the control and organisation of the Eugenics Laboratory, Pearson may have passed over too lightly in this letter

* The quotation is taken from a printed statement entitled *History of the Biometric and Galton Laboratories*, drawn up in 1920 in connection with the opening of the new building given by Sir Herbert H. Bartlett to house the Department of Applied Statistics.

the attraction which the biological sciences had for him, but I think it is true to say that however much he was fascinated in these earlier years by research into the theory of evolution and later on by the study of man, that "queen of the sciences," it was at bottom the statistical method of approach and the application of mathematical tools to the analysis of observational data which he felt it to be his main purpose to advance. The spirit of the biometric school could be described in Galton's words: "Until the phenomena of any branch of knowledge have been submitted to measurement and number it cannot assume the status and dignity of a science." Pearson may at times have been over confident of the strength of his tools, but his purpose was to demonstrate the essential need for their use in many fields. When, for example, he wrote to Weldon that one long piece of research into errors of observation* was "intended as a torpedo for the astronomical ark," the launching of this missile was in no factious spirit; no doubt, as he wrote this phrase, there was a twinkle in his eye, but beneath there was a deep conviction that the methods of mathematical statistics opened a new road for scientific investigation.

In 1904 Galton had made a gift of £1500 to the University of London for the furtherance during three years of the scientific study of Eugenics As a result, his Eugenics Record Office had been started in rooms, first at 50 and later at 88, Gower Street provided by University College, with a staff consisting of Mr Edgar Schuster as Research Fellow and later of Miss Ethel M. Elderton as his assistant. One of the first pieces of work undertaken was the compilation of a register of Able Families. The direction of the Office was entirely in Galton's hands.

Towards the end of 1906 Schuster wished to resign his appointment in order to undertake more purely biological work, and Galton, who was now 84 years old and at the time unwell, felt that the task of choosing a successor and planning a research programme was too heavy for him to undertake. He therefore decided to hand over the control of the Office to Pearson so that it might be run in contact with the Biometric Laboratory. Correspondence regarding the transfer and the objectives of what was henceforward to be termed the Francis Galton Eugenics Laboratory is set out fully in *The Life, Letters and Labours of Francis Galton* ((18) III A, pp. 296—307). David Heron, who had been attached to the Biometric Laboratory since 1905 and who for ten years was to be Pearson's leading statistical colleague, became Galton Fellow, Miss Elderton became Galton Scholar and Miss Amy Barrington, part-time Computer. The Laboratory was not transferred into rooms in the College itself until October 1907.

Pearson was somewhat hesitant in taking over this new responsibility. Besides the additional work that it threw on his shoulders, he was aware that his views on eugenic research, involving a patient collection and reduction of data, did not correspond exactly with those of Galton, who was eager for quick results and pleased with slighter contributions that would catch the public imagination. But he realised

* This interesting and perhaps little noticed piece of work was based on experiments carried out by Pearson, Lee, Yule and Macdonell between 1896 and 1900. It was published in 1901 under the title "On the Mathematical Theory of Errors of Judgment, with Special Reference to the Personal Equation" (45).

that unless he stepped into the breach this pioneer effort, which held so much promise for the future, might fail from lack of a directing hand.

And so at the beginning of 1907 we find Pearson head of the Department of Applied Mathematics, in charge of the Drawing Office for engineering students, giving evening classes in Astronomy, the director of two research laboratories and the editor of their various series of publications and of *Biometrika.* It was indeed a tremendous task which only a man possessing his power of concentration and his faculty of rapid shifting of mind from one subject to another could hope to have carried out successfully. Fortunately for his health, as one of his assistants of those days writes, "London weather saw to it that many astronomy nights must be cancelled."

It is impossible to refer here in detail to all the research work which Pearson initiated in the next few years. Apart from what was published in *Biometrika,* it was issued in three main memoir series, (i) the *Biometric Series* and (ii) the *Studies in National Deterioration,* both issued as *Drapers' Company Research Memoirs,* and (iii) the *Eugenics Laboratory Memoirs.* The *Biometric Series* contained, to start with, further papers of the series "Mathematical Contributions to the Theory of Evolution," **(43)** 1904, **(44)** 1905, **(46)** 1906, **(47)** 1907, and **(48)** 1912, those long memoirs containing mathematical theory and biological applications that the Royal Society had shown unwillingness to publish. Later it contained the three memoirs on Albinism (**(49)** 1911 and 1913), two on the "Long Bones of the English Skeleton" (**(50)** 1917 and 1919), and one on the "Sesamoids of the Knee Joint" (**(52)** 1922), which was a reprint of *Biometrika* articles **(51)**. The allocation of papers between the National Deterioration and the Eugenics Laboratory Series was probably determined by the Laboratory to which the author was attached and the funds used for publication, as much as by the subject-matter of the paper. Thus, the four memoirs on the statistics of Pulmonary Tuberculosis (**(53)** 1907, **(54)** 1908, **(55)** 1910 and **(56)** 1913) were published in the former series and the four on Alcoholism (**(57)**, **(58)**, **(59)** 1910 and **(60)** 1911) in the latter.

From the point of view of subject-matter we may usefully classify the most important of the publications of these years under three heads: (i) memoirs concerned with the collection and analysis of fundamental data regarding inheritance; (ii) memoirs in which statistical methods were used in an endeavour to throw light on important social and eugenic problems of the time, and which often involved the Laboratories in prolonged controversy; (iii) contributions mainly concerned with statistical theory.

(i) *Collection and analysis of fundamental data regarding inheritance*

Foremost under this heading comes the *Treasury of Human Inheritance,* of which Parts I and II of Volume I were published in 1909 **(61)**. It was planned on a comprehensive scale, intended to provide data in the form of pedigrees, illustrative plates and verbal descriptions for the measurement of all phases of human heredity.

"The publication of family histories," Pearson wrote in his Preface, "—whether they concern physique, abnormality, ability or achievement—whether they be new or old—is the purpose of this *Treasury*. Students of heredity find great difficulty in obtaining easy access to material bearing on human inheritance. The published material is voluminous, scattered over a wide and often very inaccessible journalistic area. The already collected although unpublished material is probably as copious but no central organ for its rapid publication in a standardised form exists at present. The Eugenics Laboratory alone possesses several hundred pedigrees of family characteristics and diseases which it is desirable to make readily accessible. Many medical men possess similar material ...

"A complete pedigree is often a work of great labour, and in its finished form is frequently a real work of art. To the many who have felt the delights of genealogical inquiry, we would say: Widen your outlook, recognise that there is something beyond names, births and deaths worthy of record, and, as it is harder to ascertain, more exciting in the pursuit. The pedigree of temperament, disease, ability, and physique which ought to replace the old nominal pedigree—if not for exhibition—at least in the family archives, is the true measure of the fitness of a stock, and the best guide to the younger members in their choice of career and alliance.

"For a publication of this kind to be successful at the present time, it should, as I have indicated above, be entirely free from controversial matter. The *Treasury of Human Inheritance* therefore contains no reference to theoretical opinions. It gives in a standardised form the pedigree of each stock."

The collection of the material was made possible through extensive collaboration with the medical profession. Some ten contributors, of whom one of the most important was William Bulloch of the London Hospital, were responsible for different sections of Volume I; the general editing and standardisation of the work was undertaken in the Eugenics Laboratory. The standard was a high one and it is easy to see Pearson's influence running throughout its 550 pages, in the care for detail, the clearness of arrangement and the striking photographic illustrations. That decision not to use the data to illustrate any theory of inheritance, but to aim at an absolutely unbiased gathering, sifting and publication of material has made the first and succeeding volumes of the *Treasury*, as its Editor had hoped, a record of great and lasting value.

There are several references to albinism in the early volumes of *Biometrika*. Both Darbishire and Schuster at Oxford had carried out at Weldon's suggestion certain experimental crossings of different races of mice, to determine how far albinism could be regarded as a Mendelian unit-character. In 1904 Weldon had contributed a note on "Albinism in Sicily and Mendel's Laws*," which led to some discussion with Bateson on the interpretation of the data. It was no doubt in order to get to the bottom of some of the questions in dispute by obtaining a much larger supply of reliable data, that Pearson in collaboration with two ophthalmologists, Edward Nettleship and C. H. Usher, commenced about 1906 to collect the material that was later published in the three volumes entitled *A Monograph on Albinism*

* *Biometrika*, III, p. 107.

in Man ((49) 1911 and 1913). The scope of this inquiry went far beyond the testing of this or that theory of inheritance. It aimed, as in the case of the *Treasury*, at putting on record a wealth of data on the subject collected from published sources and by new inquiry; combined with this was research into the character of pigmentation in the eye, the hair and the skin for both man and certain animals. The headings of the eleven chapters contained in Parts I and II provide some idea of the range which was covered: Introductory; Early Notices of the Occurrence of Albinism; Geographical Distribution of Albinism; The Albinotic Skin (Historical and Theoretical); Leucoderma; Partial Albinism; The Albinotic Eye (Man); Albinotic Hair (Man and Lower Animals); The Albinotic Eye (Lower Animals); On the Seasonal Variation of Winter White Animals; Experimental Breeding in Dogs with Reference to Albinism and Piebaldism.

In the historical chapter, as in the section on Dwarfs in the *Treasury*, Pearson's early gift for historical research found free play. Part of the secret of his immense power for creative work lay in this variety of his interests. He could turn with enjoyment and profit from algebra and arithmetic to piece together that tradition of an albino race placed sometimes in Africa, sometimes in India, sometimes in South America, which has turned up from time to time from the days of Pliny and Ptolemy; or to collect early records of albinotic or piebald negroes brought into this country as slaves. The volumes were beautifully illustrated by pedigrees, photographs, coloured plates of samples of hair and microscopic drawings of eye and hair sections, etc.

The physiological investigation made quite clear the complexity of the problem; it did not seem possible to class an individual as an albino or a non-albino, for the degree of pigmentation might vary enormously. Besides this, some portions of the body might be devoid of pigment and not others.

> "Albinism is not in our opinion," the authors wrote in the introductory chapter, "a single narrowly-defined condition, which exists or does not exist in an individual. The frequency of the individual sub-classes, and the degree of intensity even within these sub-classes, are points which require very careful consideration; it is only comparatively recently that trained observers have turned their attention to the collection of these cases of incomplete and imperfect albinism." ((49) p. 9.)

The final discussion of the material, with chapters on the vital statistics of albinism in man and the relation of albinism to other pathological states, as well as the final reduction of the statistics of heredity of albinism in man were to be issued in a fourth volume*. This has never been published though much of the material is available among Pearson's papers; it was no doubt the war, intervening in 1914, that took Pearson from a subject to which he never found time to return.

An interesting line of investigation, that had its origin in the research into albinism, was the experimental breeding of dogs. From a foundation stock of three albino Pekinese, Jack, Jill and Tong, acquired by Nettleship in 1908, some sixty

* This volume was to be called Part III of the *Albinism*; the bibliography, pedigree plates and description of pedigrees were published in 1913 as Part IV.

albino puppies had been bred by 1913. Certain of these were crossed with black Pomeranians to produce a hybrid Pompek. The object of the inquiry was to study both inheritance of coat colour and of head shape, in which the two stocks showed a fundamental difference. Four measurements of the head were taken at two or three different ages on all puppies, and skeletons and skins of certain typical animals were preserved.

A preliminary report of the experiment was published in 1913 in Part II of the *Albinism*, but the breeding was continued until the stock was finally dispersed in 1933 on Pearson's retirement*. The analysis of the very long series of results was another piece of unfinished work that he had hoped to complete when freed from College duties.

One purpose of the provisional report of 1913 was to raise two searching questions. Is it possible to explain the results of an experiment such as this by the simple Mendelian rules of dominance and segregation? If not, what are we to make of the too ready conclusions of social and eugenic reformers whose minds seem to be carried away by the fascination of a single simplified law of inheritance? Such, I think, is the substance of the questions raised in the following paragraphs:

"Of course it may be asserted that these indices [obtained from the head measurements] are very complex characters, and may be compounded of many Mendelian units. To this we must reply that albinism is of a precisely similar character, there is a very large series of characters involved in the pigmentation of different parts of the eye, the skin, the hair and the internal organs, and complete albinism of the one does not involve that of the others. Length of coat is very much of the same character, for there is an immense variety of lengths of hair on head, back, tail and legs, which vary from breed to breed. In our Pekinese the colour of the coat is of a similar character, hairs of widely different tints are found on the same dog, not only in different parts but often in the same parts, and occasionally different parts of the same hair are quite differently pigmented. If it be justifiable to use 'Jewishness' and 'Gentileness' of face as contrasted Mendelian units, one recessive to the other—notwithstanding the innumerable factors which combine to give facial expression,—we are, we hold, justified in investigating whether our relatively simple indices do or do not 'mendelise.'" ((49) pp. 483—484.)

"The problem of whether philosophical Darwinism is to disappear before a theory which provides nothing but a shuffling of old unit characters varied by the appearance of an unexplained 'fit of mutation' is not the only point at issue in breeding experiments. There is a still graver matter that we face, when we adduce evidence that all characters do not follow Mendelian rules. Mendelism is being applied wholly prematurely to anthropological and social problems in order to deduce rules as to disease and pathological states which have serious social bearing. Thus we are told that mental defect,—a wide term which covers more grades even than human albinism,—is a 'unit character' and obeys Mendelian rules; and again on the basis of Mendelian theory it is asserted that both normal and abnormal members of insane stocks may without risk to future offspring

* A further paper by K. Pearson and C. H. Usher on "Albinism in Dogs" was published in 1929 (62).

marry members of healthy stocks*. Surely, if science is to be a real help to man in assisting him in a conscious evolution, we must at least avoid spanning the crevasses in our knowledge by such snow-bridges of theory. A careful record of facts will last for ages, but theory is ever in the making or the unmaking, a mere fashion which describes more or less effectually our experience. To extrapolate from theory beyond experience in nine cases out of ten leads to failure, even to disaster when it touches social problems. In all that relates to the evolution of man and to the problems of race betterment, it is wiser to admit our present limitations than to force our data into Mendelian theory and on the basis of such rules propound sweeping racial theories and inculcate definite rules for social conduct. Even if the offspring of an albino parent be themselves normal, we cannot advise them that all is safe if they marry into normal stock; for not only is Mendelism as yet undemonstrated for human albinism, but who shall determine what is 'normal' stock, when over and over again the albino appears in the mating of two stocks which have no record of previous albinism?—Let us rather adopt the tone of the soothsayer in *Antony and Cleopatra* and when we are asked 'Is't you, Sir, that know things?' reply modestly 'In Nature's infinite book of secrecy a *little* we can read.' We await the gradual building up of more complete knowledge." ((49) p. 491.)

Now, some twenty-five years after these lines were written, perhaps the best tribute that could be paid to the time, money, energy and almost affectionate care which Pearson bestowed on the breeding of his dogs through so many years would be the reduction and interpretation of these collected data in the light of the best genetic knowledge of our day. The battle between Biometry and Mendelism is surely over.

(ii) *Research and controversy*

We must now consider a few of the publications falling under the second of the headings given on p. 166 above: memoirs in which statistical methods were used in an endeavour to throw light on important social and eugenic problems of the time. One of the most important of the inquiries of this type, to which seven of the Laboratory publications were devoted between 1907 and 1913†, was concerned with the statistics of Pulmonary Tuberculosis. In these years a great deal of money was being collected and spent in Great Britain on what was termed the Fight against Tuberculosis. With the discovery of the tubercle bacillus by Koch, the idea that infection was the determining factor held the field; popular imagination was directed towards a fight to destroy the bacillus and the conditions which were supposed to encourage its existence; advanced and infectious cases were to be isolated in sanatoria to prevent the spread of the disease; while members of tubercular stocks were told they might safely marry provided they lived with a good supply of fresh air.

Largely because no adequate data were available, the campaign was not based on any reasoned examination of figures; much of it was an appeal from the "market

* C. B. Davenport, *Heredity and Eugenics*, University of Chicago Press (1912), p. 286.

† While the present section is headed 1906–1911, it has been necessary to take some latitude in the discussion of papers; the work in most of those referred to in the following paragraphs was *initiated* before 1911 though the publication date may have sometimes been later.

place" of just that kind which might be expected to rouse K. P. in his "study." His first paper, based partly on an analysis of data from the Crossley Sanatorium, Frodsham, was published in 1907 (53). This was followed by a statistical analysis by E. G. Pope of data from the Adirondack Sanatorium in America, which Pearson completed for the press in 1908 after Pope's death (54); a year later there appeared an inquiry by Charles Goring based on the family history of 1500 criminals (63).

The main conclusions to which this work seemed to point may be summarised as follows: (i) The predisposition to tuberculosis—the tubercular diathesis—was inherited at much the same rate as other physical characters in man. (ii) The existence of a much higher correlation for tubercular diathesis between parent and child than between husband and wife was a strong argument in favour of the hereditary factor; on the pure infection theory it would be expected that husband and wife would be at least as likely to infect one another as parent would be to infect child. (iii) The correlation of diathesis between husband and wife existed, but varied from one class to another, being highest in the more educated classes. This correlation was of the same order as that found between husband and wife for a number of physical and psychical characters; it had already been described as the correlation due to assortative mating, measuring the tendency of like to marry like. In the middle classes this coefficient for tubercular diathesis seemed to be almost the same as for insanity Tubercular stocks possess certain mental characteristics and it was conceivable that members of them might be to some extent sympathetic to each other, so that there was an actual sexual selection of those likely to become tuberculous. In the same way eccentric and mentally ill-balanced stocks may have an attraction for each other.

These papers were followed in 1910 and 1913 by two joint memoirs from W. P. Elderton and S. J. Perry, (55) and (56), who carried out an actuarial investigation into the comparative mortality rates (i) of the general population, and of tuberculous patients (ii) who were, and (iii) who were not, treated in sanatoria. They were forced to conclude that there was no clear evidence of a lower mortality among the second than among the third of these classes, although it was very difficult to obtain adequate comparable material. Further, they could find no evidence to support the claims of the advocates of the tuberculin treatment.

Pearson gave an admirably clear popular account of the meaning of all these investigations in a lecture delivered at University College in March 1912, afterwards published in the *Eugenics Laboratory Lecture Series* (64). A more vigorously critical attack against dogmatic assertions by some members of the medical profession, and in particular against Newsholme's *The Prevention of Tuberculosis*, was published in 1911 in the pamphlet "The Fight against Tuberculosis and the Death-rate from Phthisis" (65), issued in *Questions of the Day and of the Fray*, a series devoted to the discussion of the more controversial topics of the hour.

If at times the controversy over tuberculosis was hot, that over the question of alcoholism raged far more fiercely. The first contribution to the subject from the

Eugenics Laboratory was published by Ethel M. Elderton and Karl Pearson in 1910 (57). Their object was to investigate from series of data from Edinburgh and Manchester whether there was evidence that the alcoholism of parents had any marked influence on the mentality and physique of the offspring as *children*; they were not at the moment concerned whether these children when adult might be likely to exhibit alcoholic or other unsatisfactory tendencies, since on this point the data used could provide no information. The difficulties of the problem and the limited scope of the investigation were set out with much care and clarity, the final conclusions may be quoted:

"To sum up then, no *marked* relation has been found between the intelligence, physique or disease of the offspring and parental alcoholism in any of the categories investigated. On the whole the balance turns as often in favour of the alcoholic as the non-alcoholic parentage. It is needless to say that we do not attribute this to the alcohol but to certain physical and possibly mental characters which appear to be associated with the tendency to alcohol. Other categories when investigated may give a different result, but we confess that our experience as to the influence of environment has now been so considerable, that we hardly believe large correlations are likely to occur.

"If, as we think, the danger of alcoholic parentage lies chiefly in the direct and cross-hereditary factors of which it is the outward or somatic mark, the problem of those who are fighting alcoholism is one with the fundamental problem of eugenics. We fear it will be long before the temperance reformer takes this to heart. He is fighting a great and in many respects a good fight, and in war all is held fair, even to a show of unjustifiable statistics. Yet the time is approaching when real knowledge must take the place of energetic but untrained philanthropy in dictating the lines of feasible social reform. We can only hope that this intrusion into the field of alcoholic inquiry will be recognised as an earnest attempt to measure the true influences of a grave social evil. Yet we have our fears, ...*."

The paper was a well-written and unbiased scientific contribution, although no doubt there was a certain challenge in the concluding paragraph and its quotation. Its publication stirred up, however, a veritable hornets' nest of critics. Cambridge economists and medical men who had written on the subject of alcohol joined with platform orators of various temperance organisations in an excited buzz of criticism and misinterpretation of the memoir. Its authors were accused of every scientific blunder and almost of social and moral delinquency. Once the battle was joined Pearson hit back with characteristic vigour. The first pamphlet of *Questions of the Day and of the Fray* series issued in 1910 (66) contained an answer to the economists who had criticised the memoir on the grounds that the populations dealt with were not fair samples of the working-class population; later in 1910 Pearson and Elderton published a joint reply to their medical critics in the *Eugenics Laboratory*

* The paper ends with a quotation from Plato's *Euthyphro* in which Socrates says (in Jowett's translation): "For a man may be thought wise; but the Athenians, I suspect, do not trouble themselves about him until he begins to impart his wisdom to others; and then for some reason or other, perhaps, as you say, from jealousy, they are angry."

Memoirs (58); and in 1911 Pearson answered further criticisms of Sir Victor Horsley and Dr Mary Sturge in another *Day and Fray* pamphlet (67).

"Dove si grida, non é vera scientia" was one of Pearson's mottoes. His line of reply was to take the statistics of various earlier writers with which his critics had attempted to confound him and to show that while in some cases they were most unreliable, in others, when properly analysed, they pointed to conclusions hardly differing from those which he and Miss Elderton had reached. It was almost too easy a task. "We have not discussed at length," the authors wrote at the conclusion of their joint reply, "all the data provided by Sir Victor Horsley and his colleagues; we have merely sampled their material to indicate how little real knowledge flows from their methods of treatment. But if occasion arises we shall go further; our illustrations are not selected, they are a random sample of the 'rebutting' evidence produced by the medical critics of our memoir."

This controversy also brought the Eugenics Laboratory into conflict with certain leading members of the newly formed Eugenics Education Society. It was a conflict which Pearson would, if possible, have avoided, partly because he knew that it pained Galton, the Founder of the first and the Honorary President of the second organisation; partly because he realised that it did no good to the reputation of Eugenics as a science. In a Foreword to the first issue of the *Eugenics Review* Galton had written: "There are two sorts of workers in every department of knowledge—those who establish a firm foundation, and those who build upon the foundation so established." Pearson doubted from the start whether these two types of workers could in fact co-operate, but he had hoped it would be possible for them to follow their own lines, leaving each other alone. The following letter, already published in *The Life, Letters and Labours of Francis Galton* ((18) III A, pp. 371—372) expresses his views clearly:

Hampstead
February 7, 1909.

My dear Francis Galton,

Thank you most heartily for your very sympathetic letter. I agree so wholly with what you say—there is need for the purely scientific research, and for propaganda. I feel that the former demands two essentials: we have got to convince not only London University but the other universities (i) that Eugenics is a Science and that our research work is of the highest type and as reliable and sober as any piece of physiological or chemical work, (ii) that we are running no hobby and have no end in view but the truth. If these things can be carried out we shall have founded a science to which statesmen and social reformers can appeal for marshalled facts. If our youthful efforts were mixed up in any way with the work of Havelock Ellis, Slaughter or Saleeby, we should kill all chance of founding Eugenics as an academic discipline. Please don't think I am narrow, or that I do not admit that these men have done or may do good work. All I say is that I could not get the help we are getting from the medical profession, from pathologists or physiologists, if we were supposed to be specially linked up with these names. Rightly or wrongly it would kill Eugenics as an academic study. All I want is to stand apart doing our scientific work, not in any way hostile to the Eugenics

Education Society, giving it any facts we can or an occasional lecture, but not being specially linked to it in any manner. For this reason I am rather sorry that X. has gone on to its Council, because it makes a link, which I think it is better for Laboratory and Society not to forge—it will hamper the freedom of both. My policy, however, with my young people is to show them my own standpoint, but in no way to control their action. Unofficially and privately I shall always be ready to aid the Society.

Yours affectionately,
KARL PEARSON.

The avoidance of open conflict became however impossible when, directly after the issue of Elderton and Pearson's first memoir on alcoholism, Mr Montague Crackanthorpe, the Chairman of the Eugenics Education Society, wrote a long and antagonistic letter to *The Times*, which included such statements as the following:

"To those, however, who are familiar with the methods of eugenic...research the Report [i.e. (57)] causes no surprise at all. It simply confirms their belief that, serviceable as biometry is in its proper sphere, it has its limitations, and that a complex problem such as that of the relation of parental alcoholism to offspring is quite beyond its ken....

"First the biometrical method is based on the 'law of averages,' which again is based on the 'theory of probabilities,' which again is based on mathematical calculations of a highly abstract order. From this it follows that in this particular problem, biometric research supplies no practical guide to the individual ..."

Galton himself felt it necessary to reply with a letter to *The Times* expressing complete dissent from the views of the Chairman of his Society, and he was almost moved to resign his honorary presidency. Putting aside the controversial aspect of the matter, the sentences from Crackanthorpe's letter that I have quoted illustrate a fact which it is always well for mathematical statisticians to bear in mind: the inevitable difficulty in putting across to the layman the sense in which the abstract theory of probability can be used as a guide to practical action. At bottom in this, as in other cases, the emotion underlying the attacks directed at Pearson and his biometric school by so wide a variety of critics was largely aroused by his claim that a mathematical technique, which they could not understand, was needed in the solution on scientific lines of the questions on which they considered themselves experts. Pearson did not seek for controversy; he knew how much time and energy it wasted. "Our policy is to work steadily away building up for the future. So long as the Mendelians do not attack us we shall leave them alone," he wrote to Mrs Weldon in 1907 on another occasion. Nevertheless his "capacity for roving into other people's preserves," coupled with his constant insistence on the need for statistically-trained minds, conveyed an implied criticism of the non-statisticians working in subjects they regarded as their own. While some sought the training that was needed, it was not surprising that many reacted in a different way. And once the dogs of war were loosed, Pearson gave blows as stoutly as they were given.

Two further memoirs were issued from the Eugenics Laboratory dealing with alcoholism ((59) 1910 and (60) 1912); they were concerned with a study of extreme

alcoholism in adults and gave particular attention to its relation to mental defect. The first paper, a joint work of Amy Barrington, Karl Pearson, and David Heron, dealt with data provided by Dr F. A. Gill, the Director of the Lancashire Inebriates Reformatory for women at Langho. The tentative conclusions reached by the authors were as follows: extreme alcoholism, like many forms of crime, was due to a want of will power and of self-control; it was therefore a consequence of the absence of mental balance, i.e. the consequence rather than the cause of mental defect as many persons had claimed. Mental defect, in many stocks at least, was an hereditary character. It followed that "segregation of the mentally defective child of both sexes was a first step in the effective treatment of both alcoholism and criminality." The authors also urged official recognition of the fact that "the prisons, the asylums and the inebriate reformatories form in combination a great national laboratory for the study of those degeneracies upon the limitation of which the welfare of society so largely depends."

The final memoir of the group (60), by David Heron, dealt with data from Inebriate Reformatories collected by Dr R. Welsh Braithwaite, the Inspector under the Inebriates Acts. From fuller data, Dr Heron reached conclusions regarding the relation of alcoholism to mental defect very similar to those of the preceding memoir. He also discussed in the light of these investigations the Government measures which had been or might be taken to deal with mental defect and inebriety.

The first paper issued from the Eugenics Laboratory on insanity had been written by Heron in 1907 (68). In this he considered the question of the inheritance of an insane diathesis, "a condition or state, which under suitable environment, the special mental or physical strain, ...may become one form or another of accepted insanity." The investigation, based partly on data obtained from an asylum at Perth, ran on similar lines to Pearson's investigation into the tubercular diathesis (53). A close correspondence was found between the intensity of inheritance of the insane and the tubercular diathesis, thus providing more evidence that tendencies to pathological defect were generally inherited in just the same manner as were physical characters. Further evidence on this point with regard to insanity was supplied by Goring in his memoir already referred to on the family history of criminals (63).

The most forceful contribution of the Eugenics Laboratory to the subject of mental deficiency belongs to a period a little later than that which I have been discussing, to the years 1913–1914, but some reference may appropriately be made to it here. The American Eugenics Record Office, after the collection and analysis of a considerable number of pedigrees and family records, had announced that there was little doubt that Mental Defect was a recessive Mendelian unit-character. On this assumption Dr C. B. Davenport, the Director of the Office, had written*: "At last it is possible to give definite advice to those about to marry, or who do not wish

* *Heredity and Eugenics*, University of Chicago Press (1912), p. 288.

to transmit their undesirable traits.... Weakness in any trait should marry strength in that trait; and strength may marry weakness." Although the danger from the social point of view of accepting this doctrine without prolonged and careful research should have been clear, the American work was regarded as of first-class importance among a wide circle of persons in this country. It was also seized on by the popular press, which spoke of the "entirely splendid work of the American Eugenics Record Office."

Pearson felt it to be essential to challenge the whole character of this work; in the first place it was necessary to show that even if mental defect was due to the absence of a Mendelian unit-character, or determiner, in the germ-plasm, Davenport's advice might lead in the long run to most undesirable consequences. A more critical study of family pedigrees showed, however, that it was not possible to fit the problem into the simple Mendelian scheme proposed; it was far more complex. Mental defect could not be regarded as a character which was either present or absent in an individual; as far as it could be measured by intelligence tests in children, there seemed to be a continuous grading with no sharp boundary whatsoever between the normal population and the population of children segregated as mentally defective. Finally, slipshod and uncritical work of this character by writers who had allowed theory to outrun knowledge was a serious offence against the infant science of Eugenics. The public, who in the long run had common sense, would put to the test such advice as "Let weakness in any trait marry strength in that trait, and strength marry weakness," would find that it failed and end by condemning wholly a science which proclaimed such absurdities.

The challenge was taken up in three pamphlets of *Questions of the Day and of the Fray,* by Heron (**69**) 1913, Pearson and Jaederholm (**70**) 1914 and Pearson (**71**) 1914. The artillery may perhaps have been unnecessarily heavy for its job, but those who have read the passages in Pearson's *Ethic of Freethought,* to which I have referred above*, will understand the deep sincerity which was associated with what he himself termed an "almost religious hatred" of error "propagated in high places." It will be well, I think, to set out here his own frank account of this aspect of the duties of a scientist, with which he prefaced his lecture on "Mendelism and the Problem of Mental Defect," delivered in February, 1914 (**71**).

"I am quite aware," he wrote, "that it is very bold for one who has had no direct experience of the mentally defective, either as a school medical officer, or as a teacher in a special school, to stand before you to-night and profess to give his opinion on the subject. But as I grow older I feel more and more the need not only for the *censores morum,* but for *censores scientiarum,* a species of watch-dogs of science, whose duty it shall be not only to insist upon honesty and logic in scientific procedure, but who shall warn the public against appearances of knowledge where we are as yet in a state of ignorance. In this age of self-advertisement, when an individual may become famous in twenty-four hours by aid of the illustrated daily press, there is quackery in science as

* Part I of this memoir, pp. 202—206.

there is quackery in medicine. And even where there is not quackery there is ignorance and dogma parading before the public as knowledge, and taking its toll from the community by a multiplicity of devices. In many ways the trained scientific mind can warn the public, even when it lacks acquaintance with specialized detail, and this is, above all, the case when the final problem turns on the interpretation of figures. To figures, in my experience, ultimate appeal is invariably made, and too often this appeal is in the inverse ratio of the power present of handling them. After all, the legitimate method in every branch of science is one and the same. The processes of observation and the material handled will differ, but the method of deducing a legitimate conclusion is common to all branches of investigation. It is summed up in the theory of logical inference, in the legitimate association of conceptions drawn from the facts observed. Unfortunately at the present time no theory of what we may term scientific logic is taught to students of science in our universities, and the result is only too patent in 50 per cent. and more of so-called scientific publications.

"I am fully aware that with so many tramps about the task of the watch-dog is by no means a pleasant one. He is thought to be quarrelsome for the fun of the fight, and writers rarely see both sides of a scientific controversy, or understand the almost religious hatred which arises in the true man of science when he sees error propagated in high places, and is told, forsooth, that he must not check this error by every means in his power for fear of hurting the feelings of Smith or Brown. There comes also a time when reasoning with error is absurd, when statements are so manifestly idle that they stand not by any force of observation behind them, but by the dead weight of authority. Then the only course open, the only thing which will kill obscurity is ridicule and sarcasm. Remember the years in which Erasmus, Reuchlin, and Agricola struggled by aid of reason alone to overthrow the scholasticism which choked all healthy growth in the mediaeval universities; then came the *Epistolae obscurorum virorum*—the everfamous letters reputed to be written by the obscure men, the scholastic theologians, one to another,—and within a couple of years the biting sarcasm of these letters of the younger humanists had freed the universities of Germany from their bondage. The renaissance had triumphed by the ridicule of obscurity, if the ground must first be cleared by the heavy artillery of scholarship and logic brought into the fight by the older humanists. To those who see the changes now taking place in the scientific world there must be a consciousness of a similar renaissance in progress. New scientific methods, new standards of logic and accuracy have fought their way to the front, and both in pure science and in medicine much of the work which may be done in the future on the old lines can only be looked upon as dogma or as quackery. The scientist and the scientific medical man have got to pass through the stage of saying *Ignoramus*, before they can safely assert that they begin to see clearly again." ((71) p. 3.)

Before concluding this account of the statistical investigations carried out during this period by Pearson or by those working under him, there are two further lines of research that must be mentioned, both of which were dealt with by the Biometric rather than the Eugenics Laboratory; they were concerned with the very different subjects of craniometry and astronomy.

The work of Quetelet and Galton in applying mathematical methods to anthropology led naturally to the application in craniometry of the biometrician's developing

statistical technique; only by the use of such methods was it possible to give precision to the characteristics of a race or to make a scientific determination of the reality of racial differences. As early as 1895 Pearson, with the help of Alice Lee and G. U. Yule, had made a series of measurements on the human skull and had calculated from these the constants of variability in man published in Pearson's *The Chances of Death* ((7) Vol. I, pp. 256—277). In the same year the first large collection of skulls and skeletons was sent to Pearson from Egypt by Flinders Petrie. This consisted of complete skeletons and skulls of over 400 members of the prehistoric Naqada race. A first investigation on this material by Ernest Warren was published in 1898*; a second report, mainly due to Cicely D. Fawcett with assistance from Alice Lee, but edited and arranged by Pearson, was published in 1902 in *Biometrika* (72). This last paper commences with a brief historical account of previous work and contains a statement of the objectives of biometric research in craniometry.

In the first place a precise definition of the characters measured was necessary. Since these characters were approximately normally distributed among the individuals of a race, any sample measured could be adequately described by the means and standard deviations and by the correlation coefficients between characters. The probable errors of these statistical constants would measure their reliability and make possible inter-racial comparisons. The correlation between two characters within a race provided little information, however, regarding the correlation between the averages of those characters in different races. While within a race the individual with a high value for a character x might tend to have a high value for a second character y and that with low x have low y, the race with high average for x might in general have a low average for y. Thus a full understanding of racial differences and of their bearing on the evolutionary history of man involved the patient accumulation of data. "A first step in this direction," the paper concluded, "should be to obtain the average values of some 40 or 50 characters in 50 to 100 races measured on some uniform plan."

From 1902 onwards a part of the energies of the Biometric Laboratory was directed to this ambitious task, as suitable series of skulls became available. Improved technique was evolved from time to time and many side issues were followed out, but this fundamental objective was kept in view. Among workers who, up to 1914, contributed in this field may be mentioned W. R. Macdonell, R. Crewdson Benington, in whose memory a Research Scholarship was founded in the Biometric Laboratory, Miss Dorothy Smith, Miss E. Y. Thomson and Miss K. Ryley. For the photography Pearson himself was largely responsible.

Much of anthropology is concerned with the study and comparison of groups. Yet it was not until Pearson's descriptive technique had been developed and tried out in many cases that it was possible to demonstrate the inadequacy of other

* "An Investigation on the Variability of the Human Skeleton. with especial reference to the Naqada Race discovered by Professor Flinders Petrie in his Explorations in Egypt." *Phil. Trans. Roy. Soc.* CLXXXIX, B (1898), pp. 135—227.

methods of approach then current, and to discover in just what sense the process of classification into groups could be carried out. Pearson was the first to insist on the necessity of obtaining large samples of skulls or bones; it was perhaps the evident impossibility of drawing sound inferences in this field from small samples that influenced his whole outlook on the general problem of small-sample theory. He also insisted on the collection of osteological material, partly because the measurements taken on the living are less accurate and partly because he felt that a profitable study of the relationships between existing groups depended of necessity on a study of their ancestral groups. In these ideas and their working out in practice lay his greatest contribution to physical anthropology.

Pearson's application of statistical methods to astronomy between 1906 and 1911 arose from his teaching of the subject in the Department of Applied Mathematics and also from his friendship with H. H. Turner of Oxford. He approached the matter with some diffidence, being "aware how badly the mere statistician may stumble in dealing with astronomical data," but he felt, I am sure rightly, that the method of correlation might provide useful exploratory tools in astronomical research. A joint paper of 1908 with Miss Winifred Gibson (**73**), following an earlier paper published by her in 1906, dealt with the inter-correlations of the stellar characters, colour, spectral class, magnitude, parallax and proper motion. A further paper of 1908 written by Pearson with assistance from Miss Julia Bell (**74**) discussed the correlation between light-range and light-maximum in various classes of double stars. This led to some discussion with H. C. Plummer—I purposely call it discussion and not controversy, for from neither side came any of the sting that accompanied the controversies in the field of eugenics referred to above—on the meaning of spurious correlation and the interpretation of numerical results*. There was an essential difference in approach that could hardly be bridged. I think we may see, too, a certain analogy with the differences that had characterised the outlooks of the biometrician and the Mendelian. The former believed that his tools, applied to mass data, could lead to the discovery of general relationships not otherwise possible to detect and he was convinced that such a discovery was a valuable preliminary to more detailed investigation regarding the individual unit. On the other hand to the geneticist, as to the astronomer, it seemed that the key to mass results could only be found by an increased knowledge of the structure of the individual model, whether this were a model of the germ cell or of the star.

(iii) *Statistical Theory*

In the '90's the biometric investigations into evolution and heredity had at times been held up because the development of statistical theory was unable to keep pace with the demands made on it by Weldon, ever full of the discovery of new and exciting problems. But Pearson's work through a decade which had opened with the theory of frequency curves and closed with that of χ^2, of contingency

* See *Monthly Notices of the Royal Astronomical Society*, LXIX (1909), pp. 128—151, 348—354, 573—585; LXX (1910), pp. 4—12, 228—229.

and of non-linear regression, had provided a technique which was competent to handle the main problems with which the biometric school was now concerned. It is not therefore perhaps surprising that in the years following the foundation of the Eugenics Laboratory Pearson made few major contributions to statistical theory. Two papers in the Drapers' Company *Biometric Series* may be mentioned.

The first, of 1906 (46), on "A Mathematical Theory of Random Migration" dealt with a problem of considerable biological importance. The immediate cause of the investigation appears to have been a problem regarding the infiltration of mosquitoes in cleared areas, put before Pearson by Major Ronald Ross. Clearly the theory had, however, many applications; the solution given was a first step which has led to further work by others*.

A second paper, (47) of 1907, was concerned with rapid methods of calculating correlation alternative to those based on the sums of squares and the product-sum. The expression for estimating the product-moment correlation coefficient from the correlation of ranks was obtained, as also an approximation to its probable error. It is interesting to note that we have here one of the first instances of the comparison of the probable (or standard) errors of two alternative sample estimates of a population frequency constant. Pearson was able to show that the product-moment coefficient had a smaller standard error than the coefficient obtained from ranks, except when the variables were uncorrelated in the population; in that case the standard errors were equal. He pointed out that for this reason, among others, the correlation coefficient should in general be calculated by the product-moment method. A situation might, however, occur from time to time in which a gain in speed would outweigh a loss in accuracy; the rank method would then be very useful.

Other theoretical papers will be found in *Biometrika*; several of these were concerned with methods of determining correlation from data classed, for one or both variables, in broad qualitative categories For example one paper, (75) 1909, gives the method of "biserial-r"; while another, (76) 1910, describes methods of calculating the correlation ratio from data classified in broad groups, known sometimes as the methods of "biserial" and "triserial-η." The calculation of these coefficients from data of this character has more than once been criticised; in particular, doubt has been thrown on the meaning of the correlation coefficient estimated by the "tetrachoric" method from the four-fold or 2×2 table, a measure of correlation which certainly played an important part in some of the eugenic and biometric investigations to which reference has been made above.

To take an example from Goring's paper on the inheritance of the diathesis of phthisis and insanity ((63) p. 9), what meaning, it may be asked, and how much weight should be attached to the correlation coefficient, $r = 0{\cdot}44$, calculated by this method from the following table? The strict interpretation of 0·44 as a correlation coefficient involves the assumption of an underlying continuous variate, the tuber-

* See for example the paper by John Brownlee, "The Mathematical Theory of Random Migration and Epidemic Distribution," *Proc. Roy. Soc. Edin.* XXXI (1910), p. 262.

cular diathesis, following in the population sampled the normal or Gaussian distribution. It is an assumption which may or may not be accepted. But even if

		Father: Tubercular	Father: Not Tubercular	Total
Children	Tubercular	63	172	235
	Not Tubercular	309	4862	5171
	Total	372	5034	5406

it is rejected out of hand, we have still to remember that the tetrachoric r is a measure of relationship, lying between 0 and 1, which provides an ordered scaling of the intensity of association—in this case between tuberculosis in father and in child.

Where more categories were available, i.e. in the general case of an $h \times k$ classification, it was possible to apply very searching tests of consistency to the four-fold and other methods of estimating correlation, by using a variety of different classifications. The agreement found was satisfactory, particularly when, a few years later, Pearson's class-index correction was available ((77) 1913). Further, when these methods were applied to actual data for which the two variables were given on a quantitative scale, close agreement between the product-moment and more approximately estimated values of r was found. Finally, the fact that the correlation coefficients of inheritance in man found from data given in these two different forms showed close agreement, while it might have been a coincidence, undoubtedly gave much weight to Pearson's belief that the broad category methods were of great practical value. Nowhere, perhaps, were they tested so critically before acceptance as within his laboratories.

Of other papers during this period we may note one on inverse probability ((78) 1907) and another which showed how the χ^2 method could be used to determine the significance of the difference between two grouped samples ((79) 1911). There were a number of short notes published in *Miscellanea* of *Biometrika*, some of them dealing with points discussed in current lectures on the theory of statistics.

I have spent some time in a description of the research work of the Biometric and Eugenics Laboratories during this period between the deaths of Weldon and Galton; they were important years, when Pearson at the height of his power was hammering home the claim for the recognition of Eugenics as a branch of science worthy of academic study. But to make the picture I have attempted to draw more complete, it is necessary to give some account of more personal interests and relationships; to look inside the walls of University College with the help of some of those who knew him at that time.

"The Professor, as you may imagine," writes Miss E. M. Elderton, "was tremendously busy, and yet he always had time to sit down and discuss an individual problem. We

did not go to his room, but he came round at least once a day to see everyone. Even after a late afternoon lecture, he was glad to see visitors from outside, whether it was Dr Bulloch from hospital or Mr Gosset with his rucksack on the way to Euston and Holyhead. ..His enthusiasm inspired us all. I remember when he was working at albinism, on two occasions I saw albinos in the street and followed them till I discovered where they lived; then the Professor with his memory knew at once whether he had the pedigrees or not.

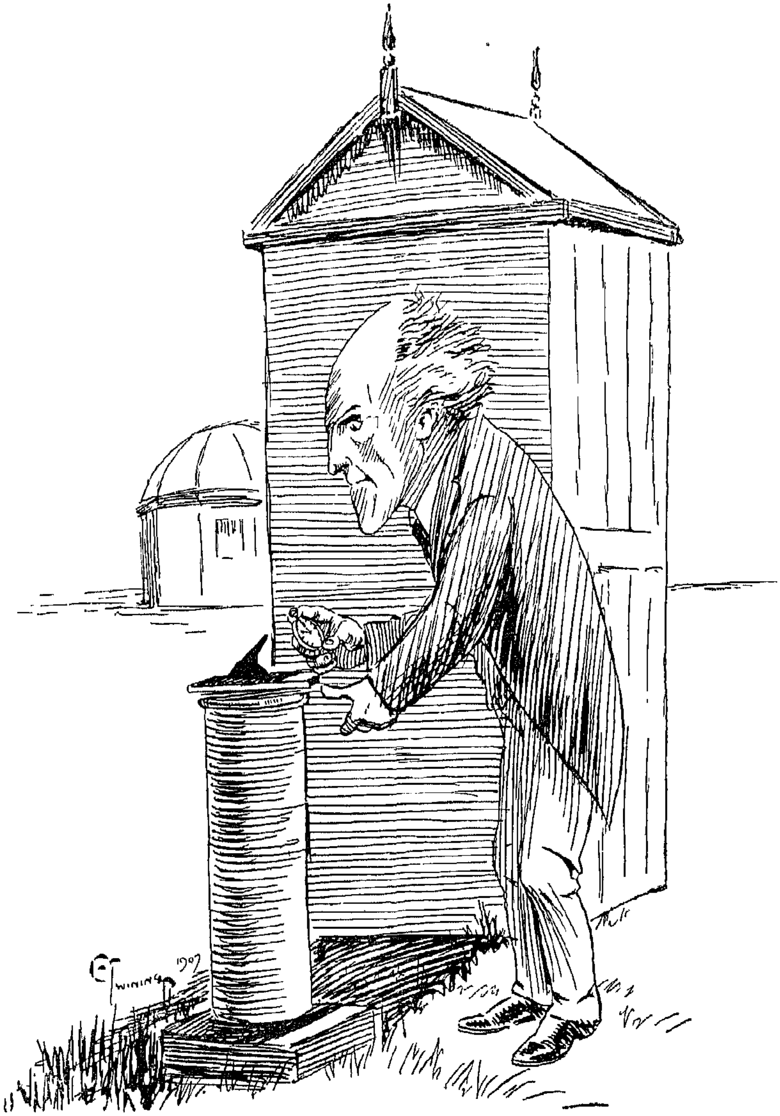
"I believe we gave the first public lectures in 1909; I remember his restlessness before a public lecture, he could not settle down to anything on the day and it used to comfort me a little to see this when I felt terribly nervous myself before one of my own lectures... In 1909 we had an evening party made possible by our move into more extensive quarters in the College. There were several 20-minute talks on the work in progress. Nettleship gave an account of the albino dogs (there were dogs there in cages); Goring talked about his criminals and I think the Professor spoke about human piebalds. We had great fun over it all and tried to make guests take away forms to fill in—family schedules, forms about numbers in the family and cousin schedules. It was the first 'party' at which I helped, and I was struck by the fact that anything short of the best would not do for the Professor; labels for example must be very tidy and printed if possible. Later I used to think that we used too much time on such things; but I wonder if we did and whether it was not important to keep up the standard in small things."

Besides the permanent research staff, there was a steady flow of postgraduate workers who came for longer or shorter periods of training in the Biometric Laboratory. Such were J. F. Tocher, Major Greenwood, Raymond Pearl, "Student" (W. S. Gosset), J. Arthur Harris, W. F. Harvey, Charles Goring, H. E. Soper, E. C. Snow and Leon Isserlis, men who have since made their names in different fields of applied statistics. Others such as W. P. Elderton and W. F. Sheppard, whose friendship with Pearson had begun several years before, were in close touch with him during this period, though never actually working in the Laboratory. Some of these have already paid a tribute in print to the inspiration they received; I shall confine myself here to quoting two further impressions of these years, given me by "Student" and by W. F. Harvey of the Indian Medical Service.

It was in July 1905 that "Student" first consulted Pearson, riding over on a bicycle from Watlington to the farm at East Ilsley in Berkshire where the latter was spending his summer vacation, in touch with Weldon at Oxford.

"I had learnt what I knew about errors of observation from Airy," "Student" writes, "and was anxious to know what allowance was to be made for the fact that a 'modulus' derived from a few observations was itself subject to error. I also wanted to know what sort of error was attached to the clumsy method which I was using to show association (difference between $S(a+b)^2$ and $S(a-b)^2$); there were also other similar questions. Pearson was able in about half an hour to put me in the way of learning the practice of nearly all the methods then in use, ready for my work in London a year later.

"I am bound to say that I did not learn very much from his lectures; I never did from anyone's and my mathematics were inadequate for the task. On the other hand I



"K. P."

Extract from "*A Collegiad*":

'Tis sure the duty of the college bard
T' extol the drapers' houses by the yard,
I should say *in* the Yard or *on* the Green,
Where, too, a modest sundial may be seen
With which to tell the time on foggy days
To guard it seemingly from Phoebus' rays
A bright-hued rabbit hutch is placed near by,
The work of Architecture's faculty.

From the *Union Magazine*, June 1907, by kind permission of the present Editors

gained a lot from his 'rounds': I remember in particular his supplying the missing link in the probable error of the mean paper—a paper for which he disclaimed any responsibility. I also learned from him how to eat seed cake, for at 5 o'clock he would always come round with a cup of tea and either a slice of seed cake or a petit-beurre biscuit, and expect us to carry on till about half past six.

"Miss Elderton and L. F. Richardson made up the rest of the lecture class. Heron was there as demonstrator and Miss Barrington as computer. Goring came while I was there, but the visit with his fellow 'criminals' was later. Crewdson Benington came then too. At that time K.P. was also running the drawing office of the engineering school and was teaching astronomy; I have a caricature of him comparing his watch with a sundial that came out in the College paper*."

W. F. Harvey came to the Biometric Laboratory in 1908 when on long furlough from India. It was at a moment when Pearson was not only beginning to enlist the services of members of the medical profession in the collection of pedigrees, but to make some of them think seriously about the meaning of probable errors and significance. In this direction he was ably assisted by Greenwood, who was pushing forward with research in medical statistics. Harvey's impressions are therefore of special interest.

"The motives," he writes, "which determine a medical man to interest himself in work outside his own profession are, perhaps, not altogether clear to him himself. In my own case they may be set down, at least partially and somewhat crudely, to the mental disturbance caused to an ardent admirer and pupil of that outstanding figure in medicine, Sir Almroth Wright, by criticism of his work†. The appearance of a destructive analysis of the figures adduced in support of the successful use of typhoid vaccine as a prophylactic measure in the South African War came as a bombshell to the believer. The correspondence which ensued introduced terms of strange application to medical argument. One heard of 'correlation coefficients,' 'significance' of differences and 'selection' A period of long furlough gave me the desired opportunity to probe further into the new instrument, which would bring a metaphorical foot rule to the measurement of causation and direct sequence in medical diagnosis, prognosis, prophylaxis and therapy.

"At the outset, and during my stay in the Biometric Laboratory, I now feel that preoccupation with mastery of details of calculation and technique obscured to some extent the full meaning and scope of the new science. The pleasure, however, of even that technical occupation was greatly increased by the opportunity I had of doing a double study, which might be described as a daily oscillation—it was a real physical one—between the spheres of influence of Sir Almroth Wright at the Inoculation Department of St Mary's Hospital, Paddington and Professor Karl Pearson at the Biometric Laboratory. That study began with attendance at 9 a.m. at University College and finished by the catching of the last train from Praed Street station back to my lodging. Both men who, as all know, have been doughty opponents took a keen

* This caricature, with an extract from a topical poem, is reproduced opposite. The "rabbit hutch" was not intended to screen the sundial but to cover a portable transit circle which was fixed onto the pedestal when required for student instruction.

† See *British Medical Journal* (1904), Pt. 2, pp. 1259, 1343, 1432, 1489, 1542, 1614, 1667, 1727 and 1775 for this controversy.

and kindly interest in my double dealing and to both of them I would now, at a much later date, tender the disciple's homage.

"These remarks are of the nature of autobiographical details. They have some significance, however, when one speaks of great men and may be the best tribute one can pay to either of them. It was inevitable that discussions should take place between pupil and masters on the subject-matter which was in dispute. I can remember well the exclamation of Professor Pearson to my suggestion that perhaps the testimony of 'experience' should not be eliminated from judgment upon a supposed cause and a supposed effect. 'Experience!' said he, 'I am always having experience thrown at my head.' Later reflection has robbed the remark of some of its original nakedness. This may be summed up by quoting the admission of another great debater of 'questions of the day and of the fray' with whom I also came in personal contact, Sir James Mackenzie, the physician. He was himself known to both of the protagonists in these lively discussions. 'Experience,' he came to admit, might be described as 'subconscious statistical arrangement of experimental data.' There we may leave the disputable subject."

The reader who cares to obtain a fuller picture of the way in which a medical man of balanced judgment could approach without bias, in those days of controversy, the relation of medicine and mathematical statistics, may turn back usefully to the paper which Harvey wrote as a result of his stay in the Biometric Laboratory *.

There is an aspect of Pearson's relations with some of his old pupils which cannot escape some comment. With many of them at some time or other during his long life he was in dispute. That intensity of purpose which carried him forward in an undeviating pursuit of what he believed was truth, brought him inevitably into conflict with the scientific views of several of his younger followers. On such occasions he felt that they were deserting a cause and he did not fully understand the effect that his strong personality had upon them. Both he and they valued independence of thought, but it was not easy for them to break free without over-emphasising what was different at the expense of so much that was common; and once that conflict of opinion had appeared on some matter which Pearson regarded as fundamental, there was a danger that those strong emotions that moved beneath the surface and were beyond his complete control would lead to a coldness on his part, an "infelicity of expression" as he had described it to Galton †, whether in spoken or written word, which seemed to make personal a difference that should have remained in the field of scientific opinion.

Nevertheless I suspect that here, as elsewhere, it took two to make a quarrel, and those of his old students who were at issue with him by the way and yet ended as his friends will agree that there were faults on both sides. There were some indeed who had the skill to differ and to win their point; that was worth while! As one writes:

"I did not always agree with K.P. Generally of course I was wrong, but if I was right and convinced him, he was always pleased about it and I went on my way feeling

* "The Opsonic Index—A Medico-statistical Enquiry," by W. F. Harvey and A. McKendrick, *Biometrika*, VII (1909), pp 64—95.

† See Part I of this memoir, p. 228.

'good all over.'...One problem, though suggested by him, was really more in my line than his; it was one of my lucky efforts when I discarded some methods of approach he had had in mind but which would not have worked and could not have given a satisfactory result. I remember being a bit anxious about it all. We had a talk when I produced arithmetical evidence; there was a pause and then he said, 'Yes, that must be right—of course you are right.' And I could have shouted for joy!"

Of friendships formed during these years, that with Charles Goring was among those which meant most to Pearson. It was a great testimony to the standing of the Biometric Laboratory that in 1909 H.M. Prison Commissioners decided to send C. B. Goring, who had been Deputy Medical Officer at Parkhurst, and two assistants, one of whom was H. E. Soper, to carry out in consultation with Pearson the statistical reduction of a long series of measurements on criminals. The observations concerned both physical and mental characters and were finally published in 1913 as a blue book, *The English Convict, A Statistical Study.*

Goring was a man of wide interests, a scholar and philosopher as well as a scientist, a born inquirer who mistrusted traditional face-values. He brought to the study of the criminal a power of careful observation and a warm humanity of a kind perhaps not often found together behind prison walls. There was something in his sympathy and understanding of his criminals, both in and out of prison, which reminded Pearson of Weldon, the naturalist, who had never been more happy than with his specimens in the field. The friendship, formed during those two and a half years when Goring was working regularly at University College, grew and developed afterwards until cut short all too soon by Goring's death from pneumonia in 1919, when battling with a prison epidemic of influenza.

I have said little of that friendship of a different kind between Pearson and Galton. Based in the first place on the admiration of a disciple for his master, it had grown more and more close in the years since Weldon's death. Testimony to this can be found in that long final chapter of the third volume of the *Life of Francis Galton.* The older man had above all provided the younger with two things, the outline of a new and powerful form of calculus and that great conception, which had so filled his later life, that "a true knowledge of natural inheritance might enable man to lift himself to a loftier level." In the last years his teaching days were over, but he could and did provide some of that wise counsel which Henry Bradshaw had supplied twenty and thirty years before. He also represented something else for Pearson; the one man who had a keen and enlightened interest in all forms of biometric work, to whom alone a Report of the work of his own Eugenics Laboratory was of capital importance.

In January 1911, in the same week in which Galton died, Pearson completed for the press Part I of the third edition of *The Grammar of Science* (**12**). It contained only the chapters on the physical branches of science, but there were included two new chapters. The first of these on "Contingency and Correlation—the Insufficiency of Causation" dealt with the author's outlook on that "category broader than causation, namely *correlation of which causation is only the limit*," to which Galton's

Natural Inheritance had twenty years before directed the young mathematician's attention. The second additional chapter on "Modern Physical Ideas" was largely contributed by E. Cunningham, at that time an Assistant Professor in the Mathematics Department at University College. Its relation to the main theme of the *Grammar* is indicated in the following quotation:

"The end of the nineteenth century, however, marks the advent of experimental knowledge requiring an entire revision of the hypotheses and theories as to the constitution of matter. In accordance with the main thesis of this work that our conceptual universe is merely the simplest logical construct into which we can gather all known perceived phenomena, the scientific mind must be prepared, as new facts of nature are brought to light, to examine whether or no they fit into the existing scheme. If they do, then the mental picture is thereby made a little more complete. If not, modification, enlargement, or even abandonment is necessary. The object of this chapter is to describe briefly the great revision that is necessitated by an unusual influx of new physical knowledge during the last twenty years." (p. 356.)

Part II of this edition, dealing with living forms, was never written; no doubt considerable addition to chapters IX, X and XI of the second edition was planned and it is clear from Pearson's Preface to Part I that he had hoped to complete this work during 1911. But new problems and responsibilities were to intervene. Whatever had been written on the biological sciences, however, in 1911, during that period of rapid development and change, must have borne a certain transitional character; nothing perhaps could have been published having the same permanent value as those nine chapters of the 1892 edition.

1911–1914

Francis Galton's death in his 89th year on January 17th, 1911, marked the end of a long and well-filled life. He had left his imprint on many branches of science, and towards the close the dominant idea of his life's work had crystallised into the conception of the linking of a new science and a new morality—for it was so that he regarded his Eugenics. It had for several years been his plan to leave the residue of his estate to the University of London for the endowment of a Professorship of Eugenics. The project had been discussed with Pearson from time to time since 1906; in particular in 1909 they had foreseen the difficulty that might occur of finding at once on Galton's death a suitable man for the post, who was young, full of energy and adequately trained in statistical method. For this reason, Galton had inserted in a codicil to his will a clause allowing the University to delay the appointment for a few years should they consider this advisable.

But Galton had another solution in mind; he saw in Pearson the ideal first holder of the Galton Chair and he realised how that appointment would release him at last "from the drudgery of teaching" mathematical and engineering students. He knew, however, that Pearson's first interest lay in the development of his training school in statistical method, the Biometric Laboratory and, fearing that Pearson would not regard the continued directorship of that Laboratory as consistent with

a loyal performance of the duties of a Professor of Eugenics, he added this final clause to his will:

"And I hereby declare it to be my wish but I do not impose it as an obligation that on the appointment of the first Professor the post shall be offered to Professor Karl Pearson and on such conditions as will give him liberty to continue his Biometric Laboratory now established at University College."

And so in the summer of 1911, after some doubts, after some negotiation with the University to ensure that there should be adequate funds not only for the salary of a professor but for the continuance of the existing organisation of the Eugenics Laboratory, Pearson relinquished the Goldsmid Chair of Applied Mathematics after twenty-seven years' tenure. He had enjoyed the work, he had learnt and he had taught, and in spite of a certain austerity of manner had won from his students all the popularity that attends a man who can hold and inspire large classes. But at the age of 54 he could have felt no serious regret at being freed to devote his whole energy to mathematical statistics, biometry and eugenics.

Many years before, Florence Nightingale had discussed with Jowett and Galton the founding of a professorship of what she had termed "applied statistics," which should be concerned with the application of statistical science to social problems*. The scheme had been dropped owing to lack of adequate funds, but the ideas discussed at that time had no doubt borne fruit in Galton's mind, and now Pearson felt that the term was an appropriate one to adopt. He was Professor of Eugenics, but the organisation of which he was in charge was concerned with a wider field; thus the Biometric Laboratory, supported by funds from the Drapers' Company, and Galton's original Eugenics Laboratory became incorporated in a new Department of Applied Statistics.

It was a research institute in the making. "There is undoubtedly work enough for two professors," Pearson wrote a few years later in a Report to the Drapers' Company, "...one to carry on the pure statistical work and biometry and the other the Galton Eugenics Laboratory. That indeed should be the goal aimed at, but it is an ideal of a distant future." The Department had yet no suitable accommodation, no adequate endowment for staff or publications, nor the funds that were needed to secure the effective attainment of the objectives which Galton had outlined in his will:

(i) *Collect materials bearing on Eugenics.*

(ii) Discuss such materials and draw conclusions.

(iii) Form a Central Office to provide information, under appropriate restrictions, to private individuals and to public authorities concerning the laws of inheritance in man and to urge the conclusions as to social conduct which follow from such laws.

(iv) Extend the knowledge of Eugenics by all or any of the following means, namely: (*a*) Professorial instruction, (*b*) Occasional publications, (*c*) Occasional public lectures, (*d*) Experimental or observational work which may throw light on eugenic problems.

* For an account of this episode, *The Life, Letters and Labours of Francis Galton*, (**18**) II, pp. 414–424, may be consulted.

In October 1911 the University issued an appeal for funds for the building and equipment of a Francis Galton Laboratory. Part of this appeal was almost at once met by the generous offer of an anonymous donor, later known to be Sir Herbert H. Bartlett, to provide a building for the combined laboratories on the Gower Street frontage of University College in continuation of the School of Architecture already under construction. The site was not altogether ideal, partly from the point of view of noise and partly because it left no possibility of later expansion, nor room for an adequate animal house for experimental breeding. But the offer was accepted; it marked a big step in the building up of that institute of which Pearson dreamed. Here is a letter written to Mrs Weldon at the end of 1912:

7, Well Road,
Hampstead, N.W.
Dec. 25, 1912.

My dear Friend,

Bed on Xmas day enables me to send a greeting to one or two old friends. Wife and bairns are at their Aunts at Highgate and I am quite peacefully convalescing from an attack of lumbago....I have read quite a lot of novels, etc. Also a couple of Greek tragedies in translation and George Meredith's letters. I am sending you the last of the index to the mice*. We shall now have to print the key and the notes on the individual mice.

I have on the whole good news as to the Laboratory. We have £3,800 from the public subscription. The anonymous donor has now offered to take the cost of building on the street front, less £3,000 to be provided by the College. This means that the donor will give about £12,000 for Eugenics and Biometry buildings and we shall spend about our whole £3,800 on equipment. We shall have a large three-storied building—not indeed in an ideal situation, i.e. on the street frontage—which is noisy and not the best for breeding work of any kind. But to have a building at all will be a great achievement. Once get this and then we can go forward to the other things I dream of!

I want to see a doubled staff with a zoologist and a medical officer and a biometric farm, such as we used to plan in the good old days! How he and I could have worked it out together, if the fates had been on our side! And now one is growing too old!...

Yours always sincerely,
KARL PEARSON.

The building itself was completed in 1914 and should have been occupied, fully equipped, by October 1915. But these plans were upset by the war.

I do not propose to discuss in any detail the research work carried out in the new Department of Applied Statistics during the three years 1911–1914. Several lines of inquiry had already been initiated in earlier years and have been described in the preceding section. To carry out the objectives set out in Galton's will, fresh supplies of the raw material for eugenic research were needed. For this purpose contact was made with the Medical Officers of Health in various large towns; data were obtained, for example, from Sheffield, Bradford, Liverpool, Glasgow and

* A reference to the completion of the reduction of Weldon's mice data.

Rochdale and in certain cases, in return, a public lecture dealing with the material supplied was given in the town by Pearson or a member of his staff. It was many years before some of this material was finally reduced and published. The following investigations undertaken, if not completed, at this period may be mentioned: (1) A co-operative study, "On the Correlation of Fertility with Social Value" (80). (2) Miss E. M. Elderton's "Report on the English Birthrate" (81). (3) An investigation based on an extensive physical, mental and medical examination of the children at the Jews' Free School in Aldgate, London, made in combination with reports from "field workers" on home conditions. The conclusions drawn from this investigation were not published until some years after the war (82). (4) An inquiry "On the Relative Value of the Factors which influence Infant Welfare" by Miss Elderton, based on data supplied by the Medical Officers of Health at Rochdale, Bradford, Blackburn, Preston and Salford. This also was not completed until more than ten years after the data were collected (83).

Besides the occasional lectures given outside London to which I have referred, a course of public evening lectures on the work of the laboratories was given at University College every winter. The following list is taken from the syllabus of one of these courses:

THE FRANCIS GALTON LABORATORY OF NATIONAL EUGENICS.

Session 1913–1914.

A course of six public lectures given on Tuesdays at 8.30 *p.m., commencing on Tuesday, February* 10*th.*

Lecture I. On the graduated character of mental defect and on the need for standardising judgments as to the grade of Feeble-mindedness which shall involve segregation. By KARL PEARSON, F.R.S., Galton Professor.

Lecture II. On some further points in connection with the fall in the Birth-rate. By ETHEL M. ELDERTON, Galton Research Fellow.

Lecture III. Infant mortality in a manufacturing town. By ALICE LEE, D.Sc., Research Lecturer at Bedford College, formerly Assistant in the Biometric Laboratory, University College.

Lecture IV. An examination of some recent studies of the inheritance factor in Insanity. By DAVID HERON, D.Sc., Assistant Director of the Galton Laboratory.

Lecture V. On the handicapping of the Firstborn. By KARL PEARSON, F.R.S., Galton Professor.

Lecture VI. On some recent misinterpretations of the problem of Nature and Nurture. By KARL PEARSON, F.R.S., Galton Professor.

Pearson's three lectures, I, V and VI, were afterwards published separately (71), (84) and (85)*.

* The dates of the lectures given in these publications do not correspond with those on the copy of the syllabus from which I have quoted; possibly the order of the lectures in the course was changed after the syllabus was issued.

Lecture V on the handicapping of the Firstborn dealt with a problem to which Pearson had referred in several earlier publications. Was it a fact that the earlier-born members of a family were to some extent inferior physically or mentally? If so, the limitation in the size of families which was spreading throughout the civilised world must by itself tend to increase general degeneracy. He had come to the conclusion that there was definite evidence of a slight "handicapping" of this character. The evidence upon which he based this conclusion had been criticised, mainly on the ground that a comparison of the orders of birth among individuals selected by virtue of their being marked in some way, e.g. by having some defect, with the orders of birth in the sibships to which they belonged, was inexact. The critics advocated a different method of comparison. Pearson was confident that he was right and his critics were equally confident that he was wrong. There we may leave the matter, since an intelligible discussion of the rival arguments would take us into detail beyond the scope of this memoir.

Lecture VI on Nature and Nurture was an emphatic restatement of the conclusion which Galton had reached long before his Laboratory had been founded, that heredity had a much greater power to determine the character of man than had environment. The lecture was illustrated by data on child welfare supplied by the Medical Officers of Health already referred to.

These lecture courses were attended by audiences both keen and critical. Social reform was in the air; a Liberal Government firmly established in Parliament seemed inclined to take an active part in social legislation; philanthropists, scientists and philosophers aired their views freely on the platform or in the press. The conception, or a misconception, of Eugenics had inevitably caught the popular attention; and while much rubbish was written and spoken, there was evidence, to be found in the pages of the weekly or quarterly reviews or sometimes in those of *The Times* and *The Morning Post*, of a thoughtful public which appreciated the value of careful and unbiased scientific inquiry on problems that must be of far-reaching social importance.

"The province of Eugenics," wrote a leader writer in *The Times* on October 7th, 1911, when supporting the appeal for funds for a Francis Galton Laboratory, "is not to yield to first impressions, but to get down to the bedrock of facts, and to arrive at correct appreciations of their value and meaning. The graver the social conditions surrounding us appear in their first aspect, the more important does it become that they should be thoroughly investigated, and that legislators and reformers should submit themselves to the guidance of knowledge in attempts to deal with them. The state of morals and of intelligence disclosed by the recent strikes, the state of health of the rising industrial population as disclosed by the medical inspections of schools are alike in showing the need for the study and the application of Eugenics, and in affording support to the appeal which we bring before our readers. It is becoming plain that the scientific investigation of the facts concerned can only be neglected by politicians who are in a hurry to introduce 'popular' reforms, and that, even with them, the neglect is more than likely to bring a Nemesis in its train."

The lectures were illustrated by many striking wall diagrams, these were at first largely planned by Pearson himself, who brought the experience acquired in the engineering drawing office to the handling of stencils for printed lettering and the arrangement of huge pedigrees which sometimes covered the whole wall of the lecture theatre. Later H E. Soper and Miss Gertrude Jones brought fresh talent into this important field of clear diagrammatic presentation.

Of Pearson's contributions to statistical theory between 1911 and 1914 the following may be mentioned:

(1) A paper of 1912 in the *Drapers' Company Biometric Series* (48), "On a Novel Method of Regarding the Association of two Variates classed solely in Alternative Categories." This contained an ingenious suggestion for calculating a measure of correlation from such data by transferring to a correlation scale the probability measure obtained from applying a χ^2 test for independence to the 2×2 table. The conception was novel; it may have originated from a search to avoid the assumption of an underlying Gaussian distribution, inherent in the tetrachoric method of calculating correlation. Abacs constructed by H. E. Soper made the computational procedure quite short, but the idea involved in the theory was not altogether simple and I do not think the method has ever been widely used.

(2) Three papers of 1913 published in *Biometrika*, Vol. IX: (*a*) On the probable errors of frequency constants (86); (*b*) On the probable error of the tetrachoric coefficient of correlation (87); (*c*) On the correction to be applied to measures of correlation calculated from data classed in broad categories, a paper to which I have referred on p. 181 above (77).

If Pearson's output of purely statistical work during these years was reduced, there was good reason. The task of writing a biography of Galton had been entrusted to him by Galton's relatives soon after the death of the latter. It could in no case have been an easy task, for to describe adequately the work, the travels, the friendships which had filled a long life of nearly 89 years must have needed much patient delving and reading. But to Pearson the undertaking was one of special significance and the standard which he set himself led him to plan out a programme which, partly it is true owing to circumstances he could not have foreseen, was not to be completed for nearly twenty years. "My object," he wrote in the Preface to Volume I, "...is to issue a volume to some extent worthy of the name of the man it bears—which may be studied hereafter by those who wish to understand him, his origin and aims...." It seemed to him in the first place peculiarly fitting to place on record some account of the ancestry of the author of *Hereditary Genius*, *English Men of Science* and *Inquiries into the Human Faculty*, all books "essentially devoted to the thesis that mental characters are inherited in the same manner and at the same rate as the physical characters." In following this course he was brought inevitably to link up Galton's ancestry with that of his first cousin, Charles Darwin,

and to attempt to trace among their forbears the many characteristics, sometimes common and sometimes different, which had marked these two great Victorian scientists. A considerable portion of the first volume of the *Life*, published in 1914, was occupied with this part of the task; it was illustrated by pedigrees and many family portraits.

"To follow step by step backwards the pedigree of one man like Francis Galton," Pearson wrote ((18) I, p. 11), "till we can go no further, but find all our lines fail us, is perhaps the most instructive lesson in history that is possible. The biographer has learnt more history, social and political, in the present inquiry than he had ever done before. One sees not only our own times linked up with great names in the past, but one feels that yeoman, squire, noble and king form a more homogeneous whole than we have hitherto appreciated with our narrow class distinctions; and we realise that the stocks which led to famous men of old may exhibit them to-day in methods more in keeping with our social ends."

It was therefore with some triumph, as artists happy in their creation, that Pearson with his collaborators, Miss Barrington and Miss Jones, must have regarded their great completed pedigree which ran back from Darwin and Galton to William the Norman, to Alfred the Saxon, to Charlemagne the Frank, to the Kings of Scotland and the Emperors of Byzantium.

If Pearson could have followed his original scheme, the biography would have been associated with an issue of Galton's collected works, which would have made the description of his many researches far easier. When, after the war, the great rise in printing costs made the plan impossible, Pearson decided that his own two later volumes of the *Life* must include a résumé of memoirs, book and articles, which had been scattered widely throughout the pages of the publications of many learned societies and scientific journals and had sometimes since become inaccessible. Only in this way, he believed, would later readers be able to appreciate what Galton had done and to pick up many suggestive lines of thought where he had dropped them.

There was another side also to this long three-volumed *Life*; Pearson enjoyed the writing of it and the contact into which it brought him with the many sides of Galton's mind. As he tells us in the Preface to the last volume written in 1930 at the close of this great labour of love:

"It may be said that a shorter and less elaborate work would have supplied all that was needful. I do not think so, and there are two aspects of the matter to which I should like to refer....I have written my account because I loved my friend and had sufficient knowledge to understand his aims and the meaning of his life for the science of the future. I have had to give up much of my time during the past twenty years to labour which lay outside my proper field, and that very fact induced me from the start to say, that if I spend my heritage in writing a biography it shall be done to satisfy myself and without regard to traditional standards, to the needs of publishers or to the tastes of the reading public. I will paint my portrait of a size and colouring to please

"Nevertheless, my head is so full of chalk-downs and clouds, and things, I can't write biometry to-night. Always, when I have been with the country, the feeling breaks out that the other folk have the best of it. The other way you live with the country and become part of it; and you dredge, or fish, or shoot something wonderful, and you describe it, and every-one sees that it is wonderful, and you all enjoy the wonder. And there is no solution, and if there were, it would not be worth the shadow of a shower flying across the country.

And this is all wicked nonsense, and I am going to bed."

W. F. R. WELDON
Letter, April, 1903

DESSERT IN COMMON ROOM

TOASTS

1. IN PIOUS MEMORY OF SIR FRANCIS GALTON
2. IN REMEMBRANCE OF ALL BENEFACTORS
3. IN MEMORY OF THE BIOMETRIC DEAD
4. THE GUESTS
5. THE POSTGRADUATE WORKERS
6. THE PROSPERITY OF THE LABORATORIES

COFFEE IN THE MUSEUM

W. F. R. WELDON, 1860–1906
From the bust by H. R. HOPE PINKER

From the toast-card of the Seventh Anniversary Dinner, Galton and Biometric Laboratories. *February* 22, 1926.

myself, and disregard at each stage circulation, sale or profit. Biography is thankless work, but at least one can get delight in writing it, if one writes exactly as one chooses and without regard to the outside world! In the process one will learn to know—as intimately as any human being can know another—a personality not one's own, that is the joy of spending years over a biography where there is a wealth of material touching the mental output, the character and even the physical appearance of the subject."

In the three volumes of the *Life* there were reproduced from photographs, sketches and paintings nearly forty portraits of Galton. To Pearson one of the highest functions of the painter or sculptor was to catch and hold for later generations some part of the personality of the great men and women of the day. He felt that there was something which the artist with his brush or chisel could achieve that lay beyond the power of the writer with his pen. It is in this idea that I think we may trace the origin of Pearson's friendship and admiration for H. R. Hope-Pinker, the sculptor. The full-size statue of Darwin in the Museum at Oxford, a photograph of which formed the frontispiece to Volume I of *Biometrika*, and which still figures on the standard buckram binding-cases of this journal, had been carried out by Hope-Pinker. After Weldon's death, the work of modelling and casting a bust of him in bronze, also for Oxford, was entrusted to this same artist, who had perforce to work only from photographs.

"Don't mind if it is not a great portrait—it hardly can be—," Pearson wrote to Mrs Weldon, "but if he gives a work of art, which portrays a man of intellectual strength and keen mind, then be happy. It will associate your husband's memory with an ideal for future generations, who won't care much what any of us were really like in the flesh. The portrait will not live, but the ideal man of science embodied in a real work of art will."

There is a letter written to his son at school in 1912 which gives a picture of the lighter side of some of Pearson's many activities, the Galton *Life*, sculpture, dogs, a Royal Society soirée and the investigation into the temperatures of school children:

7, *Well Road*,
Hampstead, *N.W.*
May 25, 1912.

My dear old Boy,

I don't often bother you with a letter, and I have no paternal advice to give, but I thought I might write one of Mother's three weekly letters. Old Samuel Galton found one of his sons in trouble over something, and said to him: "Tell your friend Sammy all about it, and he'll say never a word of it to your Father"—and I think the old Quaker's division of the Father into business and gossip was rather good. Well this is gossip! Of course dogs come first; they have had no Saturday walks with me because I have been going to Hope-Pinker, who asked very nicely to have a study of my head for some bust he is making of Roger Bacon!! Now Matthew Paris tells us that a "quidem Rogerus Baconus, clericus de curia," was jocund and merry and fat—and the said sculptor has got a veritable Cassius both in his model and his actual study! But it

is interesting to see a big man at work, and envy his powers; there was such a shade of difference only when after working two hours on the first day he put the callipers to all parts of my head and tested on the clay itself. His appreciation of lengths *in space* must exceed that of a good draftsman on a board; we usually get to a millimetre or two in the latter before the scale is applied. E.g.  is I should say about 3·2 cm. Well, Hope-Pinker does that sort of thing *in space*.

Mr Usher kindly offered me Donald dhu, but I have nowhere to put or keep him, otherwise I should have accepted him straight off. He would run so nicely with a white dog. Perhaps some day we shall have place for him.

About Mr Croft, please give him a message from me*: (i) I suppose he saw me upstairs, but I did not see him, i.e. I may have seen him but did not recognise him. I was only upstairs in a hurry as I could not leave my dogs. (ii) Why did he not come down and talk to me and the dogs? We were all pining to see people and did not get very many as we were in an out of the way corner, behind the cloakroom. Don't forget to tell him this, and I should immensely have liked a talk with him, of course not-at-all about you!

Have the temperatures been rising in Winchester lately? I wonder if I shall get a bad name or whether you will all consent to the torture of being taken in view of the aim: i.e. to find out whether the children of the poor are really in such a bad condition as our temperature observations in their schools seem to show. I must not now write more, or I shall get no work done this morning.

Ever your affectionate Father,

K. P.

The Weldon bust had been one item of a memorial scheme initiated shortly after Weldon's death; the greater part of the funds collected were handed over to the University of Oxford to found a Weldon Memorial Prize in Biometry which is now awarded every three years. The first award was made in 1912 to Pearson, but he would not accept the honour and the reasons which he gave express characteristically what he felt on the subject of prizes and medals:

"I do fully appreciate . the desire of the Electors, but you know that I knew Weldon very closely and can still feel what he would think and say. The Darwin medal came to me when I was relatively young and it encouraged me as a young man and made me feel that medals and prizes might be helpful to young men, directing their energies and telling them that they were appreciated. The R. S. usually gives its medals to old men, whose reputations are already made, it gives them momentary pleasure and saves the R. S. much trouble in selection, but from the standpoint of science the medals are idle. Now what I have written down is what W. F. R. W. would have said, and how he acted when he proposed my name for the Darwin Medal. The Weldon Medal must go to encourage young men if it is to be fruitful to science. I am old now, and medals or no, I shall go on with the little work I still can in Biometry to the end, but a young man or two may be preserved for work in that direction ..."

* This is a reference to a Royal Society soirée, at which the Galton Laboratory had an exhibit of albino and coloured dogs, and which was attended by W. B. Croft, the Winchester Physics master.

The year 1914 saw the completion of a long-planned project in the issue of *Tables for Statisticians and Biometricians* (88). The tables had been calculated over many years and largely published, as opportunity offered, in *Biometrika*; as they were printed, they were also moulded, in order that stereos might be taken for reproduction.

"From the beginning of this work in 1901," Pearson wrote in the Preface, "when the first of these Tables was published and moulded, I have had one end in view, the publication, as funds would permit, of as full a series of Tables as possible. It is needless to say that no anticipation of profit was ever made, the contributors worked for the sake of science, and the aim was to provide what was possible at the lowest rate we could. The issue may appear to many as even now costly; let me assure those inclined to cavil, that to pay its way with our existing public double or treble the present price would not have availed; we are able to publish because of the direct aid provided by initial publication in *Biometrika* and by direct assistance from the Drapers' Company Grant."

It was in this spirit that the whole of the publications of the Biometric and Eugenics Laboratories were issued during Pearson's long directorship. There were never profits and the cost, to the reader, of tables, pamphlets or memoirs, was worked out on a basis that would do little more than ultimately repay the cost of printing. The Tables of 1914 were regarded as a first contribution to a more complete series; among other things Tables of the Incomplete Gamma and Beta Functions were already projected. The Preface gave an opportunity of thanking the many loyal friends and colleagues who had ground their Brunsvigas to the appointed end—W. F. Sheppard, W. P. Elderton, Alice Lee, P. F. Everitt, Julia Bell, Winifred Gibson, A. Rhind, H. E. Soper and others. Warm thanks, too, were offered to the staff of the Cambridge University Press. "To those who have had experience of numerical tables prepared elsewhere," the Editor could write, "the excellence of the Cambridge first proof of columns of figures is a joy, which deserves the fullest acknowledgement."

And so in the early summer of 1914 the auspices for the future of biometry and eugenics were good. However much he might himself doubt his power of continuing much longer at work, the Head of the Department of Applied Statistics, even at 57, seemed to his friends as young and full of vigour as ever. A spacious new building was nearing completion across the College quadrangle; funds for its equipment were in the bank, and plans for its Anthropometric Laboratory and Museum were being eagerly discussed. The Laboratory publications, whether in lighter or heavier vein, were purchased in increasing numbers, the sales bringing in some £250 a year; *Biometrika* was at last on a sound footing and paying its way. Courses of public lectures were well attended; though sometimes hidden behind a screen of controversy and of journalistic popularisation of the concept of eugenics, a growing body of opinion was learning to appreciate the value of statistical method. And then on those halcyon days, which to many of us now form only a dim background to our school-days, burst the thundercloud of European war, destroyer of so many hopes.

1914–1920

With Volume I of the *Life of Galton* and the *Tables for Statisticians* off his hands and through the press, Pearson had decided to break his usual summer vacation habits. Perhaps it was the discussion of Galton's *Wanderjahre* in the *Life* that caused him to revisit the Black Forest in search of the landmarks and faces that he had known so well in the '80's. Whatever the cause, early July found him established with his wife at his old haunt, the *Gasthaus zum Ochsen* at Saig, near Lenzkirch.

"Into this small society," he wrote*, "we settled down and thought neither of war nor of racial feeling. We shared our newspapers, collected scarce flowers with the German boys, or discussed where bilberries or wild strawberries were to be found. We reported at meal times how many wild deer we had startled, or planned mild expeditions to neighbouring villages to drink our coffee or test their cake. We discussed food, perhaps a trifle too insistently, but perhaps not so continuously as we should have done thirty years ago. The *Bürgermeister* was a well-to-do peasant, and he, and an old friend the *Bierwirth*—with whose father I had years ago a common interest in the pursuit of trout—would spend time in watching the upward progress of their new *Rathhaus*.... There was confidence and friendship on all hands; the local banker cashed readily my circular notes, and we waited the days when our family from college and school would join us.

"Then suddenly fell upon us the blow of Tuesday, July 28th—the declaration of war by Austria against Servia—which, to my mind, left no loophole for diplomacy to save the situation...."

After describing the resentment of the local peasantry and many of the visitors when the *Bürgermeister* announced that they were in *Kriegszustand*, he went on to tell of a broken journey home by train along the Rhine; of an increasing popular war-fever in the industrial Rhineland; of the insistent question in the train, "*Was macht England?*" of the crowd which rushed down the platform at Crefeld shouting "*Ein Russe, ein Russe*"; of crossing the frontier into Holland with luggage on a barrow; and of final arrival in London, tired but without loss of property, on August 4th, the day that Great Britain declared war. It was a strange chance that had sent him to Germany after twenty-five years, in this year of all years!

The new situation was promptly faced by the members of the staff of the two Laboratories; the right course, they agreed, was to give up their immediate programme of research and to place their calculating powers at the disposal of some Government Department. Within a few days, Pearson had arranged with the Board of Trade to set in train and carry on a piece of work for the Labour Exchanges and Unemployment Insurance Department. The scheme was to provide and keep up to date charts showing the state of unemployment among insured and uninsured workers, male and female, for each town in Great Britain of over 20,000 inhabitants.

* The quotation is taken from an article "Germany in the Eye of an Onlooker," contributed to *The New Statesman* of 22 August 1914.

These charts were required by the Department and various Relief Committees. At the initial stage something like 50,000 percentages had to be calculated from data supplied by the Labour Exchanges and 2000 curves drawn on solid paper and then transferred to tracing paper for blue-printing. The staff, to whom was added a motley crew of some ten or twelve volunteers—engineering demonstrators, teachers, undergraduates—set about the big task with a will, and by mid-September the work was settling down into the routine form in which it continued until the following June.

Such was the first piece of war work; there were many others undertaken in the next four years, culminating in the big programme of gunnery computation. I shall let the account of this work stand in Pearson's own words, in the Report he made in February 1918 to the Worshipful Company of Drapers on the "War Work of the Biometric Laboratory." This Report is given below in Appendix III. It was written, as were other similar Reports, not only as a statement of work done, but to persuade the Drapers' Company to renew their grant*.

The programme of lectures in the Department of Applied Statistics for the Session 1914–15 had included a new departure in the shape of an introductory evening course to be given in conjunction with the Department of Zoology. In the first term Dr C. H. O'Donoghue of the latter Department was to discuss "The Biological Basis of Heredity"; in the second and third terms Pearson was to deal with "The Statistical Basis of Eugenic Theory" and "The Facts and Theories of Heredity." The syllabus of Pearson's first course, with its "Games of Chance, Tossing, Lotteries....The beginnings of Statistics; Graunt, Petty....The Universe regarded as a system of correlated organic and inorganic factors; the category of correlation as replacing the conception of causation..." is reminiscent of the syllabuses of the Gresham Lectures, given twenty years before. But the scheme, alas, was still-born

"About eleven persons besides the Laboratory staff attended O'Donoghue's first lecture, 'a most excellent discourse' as Pepys would say, and of these two or three were merely first nighters and the remainder chiefly ladies diligently knitting for the soldiers. There was such a contrast to our usual throng at public lectures and so little public desire for eugenic instruction at this time, that the lectures were postponed *sine die.*"

Now, long after, with a knowledge of Pearson's fifteen years of vigorous post-war directorship, it is perhaps difficult for us to realise the most personal aspect of the strain and depression which for him accompanied those war years. Within him there was a continual struggle between loyalty to his country and its cause, which he had at heart, and loyalty to the objectives of his own life's work. A younger man could have thrown biometry and eugenics from his mind for the period of the war, confident in his own power to pick up the threads where they had been dropped. But Pearson, weary after a ten-hour day on bomb or shell trajectories, could feel little confidence in the future. Would he be there to set going that

* His dislike for this recurrent task of report writing is expressed in a comment on one of these occasions, "horribly self-laudatory, how I hate the doing of this sort of thing!"

research institute of his dreams in the building now occupied by convalescent soldiers, or must all the threads that led back to Weldon and to Galton be broken? With the old trained staff scattered and attracted into other fields; with *Biometrika* perhaps closed down for lack of funds; with a new Galton Professor in charge, might not the war gap cut finally across the line of tradition and bring to nothing all those years of effort?

In the first year of war the strain was increased by uncertainty as to the future of the new building; it was first stated that the War Office wanted to make use of it as an extension of the space allotted to the wounded in University College Hospital; then the scheme was dropped and plans for its equipment as the Galton Laboratory were brought up again, only to be dropped in turn when the War Office returned to its original project. In the planning and counter-planning which resulted, the members of the College Committee and of the Galton Committee, with the exception of Galton's nephew and executor, Edward Wheler-Galton, each no doubt absorbed in his own troubles, did not understand the struggle that was tearing at Pearson's heart.

With his own staff, too, it was at times difficult. In keeping them together as a trained unit he was following a sound practical course, as well as thinking of the future. But he could not provide the salaries that were being paid elsewhere for other Government work, and in the restlessness of those days the routine of computing and drawing charts of unemployment or of imports was not felt to be "real war work." The calculation of shell trajectories was more nearly the genuine thing, but that was not undertaken until later on. And so, while he gave his blessing to each leaving member of his staff, the loss left him with some feeling of bitterness, for old colleagues seemed to be deserting the cause which was such an essential part of his life.

The following extract from a letter to a friend was written on January 1st, 1916, after a relatively free period between two spells of war work:

"All last session we did nothing but Government work, and this term and the vacation I have been entirely occupied with individual people's work. I had to get endless papers ready; the younger generation have to push their way forward, and publication of their work is immediate and essential to them. I have nobody now, but myself, who can even whip a paper into decent form for press. In the part of *Biometrika* just out are two memoirs which I wrote up almost entirely and for which I did all the photographs. I don't complain of this and I don't want other people to know, but they mean all the work I have done in four months and I want four months for the mouse paper* and six months for the second volume of Galton—and meanwhile any work of my own gets pushed to the Greek Kalends! Meanwhile too I have volunteered to put the whole Laboratory staff again at the disposal of a Government Department and do not know how soon we may be called off all work again.

"People are writing to me about Vol. II of the Galton, as if I were a criminal for not having issued it! But they do not stay to ask what it cost in time or money taken from

* This refers to the completion of Weldon's mice work.

other things. One man wrote about 'its outrageous price,' when it would not have paid for itself had all copies been sold, and when practically the book was a failure owing to the outbreak of war. Even *Biometrika*, which seemed safely in port after stormy years, has had a frightfully bad time during this war year, but I am told that other scientific journals are in worse case*."

Among the relaxations that Pearson found in these dark years, there were two that may be mentioned; the quiet atmosphere of a cottage in the country and the stimulus which he could get by turning from guns and bombs to research into the theory of statistics. Early in 1916 he decided to look out for a week-end cottage which would provide his family and himself with a breath of fresh air in the middle of strenuous occupations. An advertisement in *The Observer* sent him down one snowy Sunday to climb Leith Hill in Surrey and to find at Coldharbour the little house which was to be his, first as tenant and then as owner, for the next twenty years.

The position of the Old Schoolhouse, as it was called, under the crown of the hill with the commons and pine woods behind and in front the wide view over the weald of Sussex to the South Downs, grew on him more and more with the years. In those first days they were cutting the larch woods for pit-props; the smell of the timber reminded him of his loved Black Forest; there was a camp of German prisoners engaged on part of the work and a chat with these men helped a little to throw into perspective that bitter struggle going on across the Channel. The country, too, was made familiar from the word-pictures of George Meredith; Box Hill, the yew trees, the wild cherry, the seed of willow herb....Though born in London, Pearson had in him a strain of the essential countryman, inherited or acquired from his father; a love of the pure air, of watching the habits of birds, the growth of trees and flowers, the changes in the clouds and sky. Ten years before he would have linked this bent with some line of biometric research. Now he felt too tired and pressed; but it gave him immense joy to scent the pine woods as he climbed the hill from the station on a Friday or Saturday night after a full week's work in Gower Street, or, on a short Easter holiday, to prepare the ground in the small garden plot and sow his lines of peas and beans and potatoes.

The death of W. R. Macdonell in 1916 deprived Pearson of yet another friend; Macdonell had formed one of the little group who had provided the original guarantee fund for *Biometrika* and he had acted until the end as an assistant editor. "Few abler proof readers," Pearson said, "can be found than a Scotsman trained in Oxford, especially if he has graduated in science, and tempered his science with modern European literature as a hobby." Macdonell's patient labour, his sound advice and lovable disposition had assured him a welcome place in the Biometric Laboratory during the three or four years which he spent there between a business career in India and London and his retirement to Aberdeen. There he had held

* A little later several warm friends of biometry, among whom Mrs W. F. R. Weldon was a most generous contributor, came to the rescue of this journal and helped it over a very difficult war and post-war period.

a part-time lectureship in biometry at the University, a post since filled by another Scotsman of strong personality, J. F. Tocher.

In the war years the issue of *Biometrika* was inevitably slowed down, but the quality was well maintained. "My pride as long as I can," Pearson wrote, "is to show no sign of war in the journal, i.e. that it should be as well done as ever." For himself, to be absorbed in algebra and searching for some new statistical result, was to forget for a few hours the war-time troubles, and as no large scale research in applied statistics was possible, it was natural that he turned to mathematics. As a result several rather interesting contributions to statistical theory were completed in these years. In 1914, R. A. Fisher had written his important paper on the "Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population*." This paper verified mathematically "Student's" predictions of 1908 regarding the distribution of a standard deviation† and of a correlation coefficient, r, in samples from uncorrelated material‡; it also confirmed the substantial accuracy of certain approximations of Soper's for the general distribution of the correlation coefficient§. Besides deducing the exact distribution of r in samples from a normal bivariate population and suggesting that a mathematical transformation would be useful in the case of high correlations, the paper illustrated the value of that device which has played so fundamental a part in much subsequent work, the representation of a sample by a point in multiple space. Pearson, grasping the importance of some if not all the aspects of this paper, with characteristic eagerness was already early in 1915 planning to put theory into numbers. The standard deviation distribution was discussed in an Editorial note (89) following Fisher's paper. Very considerable algebraic treatment and computation were however needed to explore the nature of the distribution of r; the resulting paper, a long piece of co-operative work which had occupied the laboratory staff in free intervals between war work, appeared in May 1917 (90).

The sampling distribution of r, leading to curves of remarkably different shapes, made a strong appeal to Pearson's imagination. In its variety of form it had almost the elasticity of his own system of frequency curves; his pleasure in visual representation led him to plan the construction of the models still preserved in the Department of Statistics at University College, photographs of which were reproduced in the co-operative paper. To one for whom the conception of correlation had played so important a part, Fisher's theoretical rounding off of many earlier partial solutions might have had even more significance. But if controversy and acute differences of opinion were to follow, history, which, with time, must roll out all bitterness of conflicting personalities, will find here one of the striking links between Pearson's and Fisher's work.

Problems connected with mean-square contingency and the use of χ^2 were much in Pearson's mind at this time. Looking back in the light of new ideas which have

* *Biometrika*, x, pp. 507—521; the Part containing this paper was published in May 1915.

† *Ibid.* VI (1908), pp. 1—25. ‡ *Ibid* pp. 302—310.

§ *Ibid.* IX (1913), pp. 91—115.

been evolved in recent years, we can see how he had pushed his methods of analysis forward to that stage of complexity where it was essential that some clarifying concept should be found to bind the whole into a simple scheme. In these papers he was still building, providing new results, many of which have been incorporated into the statistical theory of to-day, but the bricks did not altogether fit, there was some confusion in the pattern. What perhaps was needed most of all was a shift in outlook on the function of probability theory which would throw many results, so far disconnected, into a new perspective. The shift was not a large one, but it needed younger heads A new generation of statisticians, turning fresh minds to the problem, could make the change almost unconsciously, and because they did not realise that Pearson's view-point was different, looking back to his work might say in surprise, here he was "wrong," there he did not "understand." So no doubt, too, will a younger generation of enthusiasts, twenty-five years hence, treat the statisticians of to-day*!

Two papers, (91) 1915 and (92) 1916, the latter written jointly with Andrew W. Young, dealt with the standard error of $\phi^2=\chi^2/N$, the mean-square contingency of a two-way table; a correction of an error in the final formula was given in the paper entitled *Peccavimus!* (93) 1919. As a special case, the corrected formula provides the exact standard error of χ^2 in the case where the expected frequencies are known population values; but there appears to be a certain confusion in the basis of the work, and I am inclined to think that, from the practical point of view, the standard error obtained was not really what the writers needed to find†.

A far more interesting paper was that on Multiple and Partial Contingency, (94) 1916. Suppose that the individuals in a population fall into one or other of k categories, and that n_t and m_t are the observed and expected frequencies in the tth category ($t=1, \dots k$) found in a random sample of $N=\Sigma\,(n_t)=\Sigma\,(m_t)$ individuals Then Pearson considered the sampling distribution of

$$\chi^2=\sum_{t=1}^{k}\left(\frac{n_t-m_t}{\sqrt{m_t}}\right)^2=\sum_{t=1}^{k}(X_t^2), \tag{i}$$

where q linear relations of the form

$$h_{s1}X_1+h_{s2}X_2+\dots+h_{sk}X_k=H_s\ (s=1,\dots q) \tag{ii}$$

hold among the k values of X_t, the h_{st} and H_s being known constants. Making use of the geometry of multiple space, he showed how the distribution of χ^2 would depend on what we should now term the number of degrees of freedom among the X_t's, namely $k-q$. The chance that in random sampling χ^2 should exceed any given value, say χ_0^2, could be obtained by entering Elderton's Table with $n'=k-q+1$

* I find the following sentence in a letter of 1898 from Pearson to Weldon: "I was a good deal drawn by Galton's letter for it seemed to me that he was still hopelessly at sea with regard to the theory of regression, and if *he* did not follow the bearing of my January paper on the law of ancestral heredity, who in the world can I expect to?" But such is the accompaniment of progress, the young men step on from where the old men halt.

† The standard error of the expression, ϕ^2, that would ordinarily be calculated in practice was obtained several years later by T. Kondo, *Biometrika*, xxi (1929), pp. 376—428.

and $\chi^2 = \chi_0^2 - H^2$, where H^2 was a function of the constants h_{st} and H_s which would vanish in the common case where $H_s = 0$ $(s = 1, \dots q)$.

This paper seems to me to contain the basis of a large part of our present χ^2 theory. Two further steps, however, were needed before that theory could be reached. In the first place it must be realised that this general solution involving the linear conditions (ii) could be applied, after some adjustment and approximation, to provide a solution in the case where the quantities m_t of (i) were not known *a priori*, but must be estimated from the sample. Secondly, a slightly different, perhaps a broader conception of the meaning of a probability measure in connection with a statistical test must be evolved. Neither of these steps was taken by Pearson nor, I think, at a later date were they ever altogether accepted by him as being justifiable.

Another paper of interest published in the same Part of *Biometrika* was that on the goodness of fit of regression curves (**95**), a problem for which the immediate suggestion was no doubt a paper by E. Slutsky on the same subject*. Pearson's work, which started without assuming that the array distributions were either normal or had a common standard deviation, led him by an arduous route towards the goal. His study of actual distributions had convinced him that when regression was not linear these conditions were rarely satisfied, and I fancy that he was too honest and too self-consistent to make the assumption which has opened a gate to a great field for further development of method. It is not that later work has been in any sense dishonest; it has followed a rather different line of attack, first deriving a technique based on certain simplifying assumptions and then showing that this technique will still be valid even when the initial conditions are not rigorously satisfied. In any case we can only admire the courage with which Pearson plunged towards an approximate solution through the heavy algebra that was required to deal with the variations in means, in standard deviations and in array totals. It was perhaps the obvious need to attempt some simpler solution, even if based on less general conditions, that led to R. A. Fisher's paper on the goodness of fit of regression curves, published six years later†.

Another of Pearson's contributions of 1916 was the last of "The Mathematical Contributions to the Theory of Evolution," No. XIX, a paper dealing with certain special types of his system of frequency curves. This was published, like the first two papers on the subject, in the *Philosophical Transactions of the Royal Society* (**23**).

During 1917 and 1918, beyond the completion of the co-operative paper on the correlation coefficient already referred to, Pearson published only two short statistical papers. It was the period of intensive computation on the gunnery programme, when he had working under him a staff of twenty persons, of varied

* "On the criterion of goodness of fit of the regression lines and on the best method of fitting them to the data," *Jour. Roy. Stat. Soc.* LXXVII (1914), pp 78—84.

† "The goodness of fit of regression formulae, and the distribution of regression coefficients," *Ibid.* LXXXV (1922), pp. 597—612.

training and ability, for each of whom he must always have ready an appropriate job. "I get no spare moments now; it is anti-aircraft guns all day long, and for me most Sundays also," he wrote in June 1917.

The strain of this work began at last to tell and in April 1918, after fifteen months, Pearson felt that he must arrange for the group to be taken over by the Ministry of Munitions. For the last summer of the war only he and Miss Elderton remained in an empty laboratory, finishing up one or two special gunnery problems. Then at last came a month's holiday at Coldharbour, when he returned with delight to biometric work and to his garden, and so, as the following letter shows, shook off the insistent memory of the guns.

The Old Schoolhouse,
Coldharbour,
Nr. Dorking.
August 4, 1918.

My dear Miss Elderton,

I wonder how you are getting on with your holiday, I hope as excellently as I hitherto have done. First, I have had a real mental holiday—all the refreshment that comes from again taking up one of my old friends, the femur paper*. I have finished the hardest part of the task, chapter VII, on the individual characters; it was the longest chapter, perhaps, 80 pp. of printed matter with at least 70 tables of comparative racial material and it ought to form a good illustration for biometric work on these lines in the future. I only hope I have made no bad "howlers" to discredit it. Now I have commenced the last chapter on Primogenial Man and am about 1/3 through it so that I have great hopes of finishing the whole thing before my return. It would be so delightful to get it out and show our friends (? foes) that the Laboratory is not dead!

To this I give 4 hours in the mornings and 3 at other times in the day and the mental change is really delightful. I don't dream guns any longer! Then after midday meal I turn very vigorously to gardening and try to console myself with Kipling's "Better men than you have started on their lives, By weeding gravel paths with broken dinner knives." I expect they had better lumbar muscles, however. I have got about 1/3 of the garden weeded and hoed but it is an endless task. The rambler roses are perfect just now and the evening primroses, but there is a lack of the ordinary smaller fry. Sweet peas were a failure and so largely were Shirley poppies. We have found perpetual spinach a great success and if you have not tried it, it is well worth doing....The potatoes and turnips are good. Peas and parsnips a hopeless failure and no signs of vegetable marrows yet. We have made a good deal of jam, but currants and gooseberries were not numerous. On the whole we have done well without any real gardening help. Then there is always exercising of the dogs to be considered. We have Hans and Meg and Dinah here, but so far have not succeeded in adding to our pedigree.

Do you remember Miss X. of New Zealand? I had a long letter from her this morning on a *printed* form in which Sir Francis and I both appear as founders of a new New Zealand Eugenics Society established this year. She asks me for a cablegram to say I approve and encloses an order for *three* words. I find the least I can say is *six*

* See p. 205 below for reference to this work.

words, i.e. "X., Hastings, New Zealand, Pearson disapproves." But why should I have to pay 4/- extra, ..? It is a funny world....

May I have a line to say what you are doing? I hope you are resting.

Yours very sincerely,

KARL PEARSON.

Back at College on September 1st, he was ready to start training two or three members of a new staff. To what work he would have turned then had the war continued, I do not know. At the hour when the Armistice was signed, on November 11th, 1918, he was giving a first lesson on a Brunsviga calculating machine to a wounded New Zealander who has since become one of the leading experts on scientific computing.

The year 1919 was for all of us a year for stock-taking and adjustment. After the sudden release of tension in November 1918, we began to look round and wonder how we stood; could we go back to where we had left off in 1914? No, that was out of the question! What was lost? What remained? For Pearson there had been no war casualties of close friends or relatives, but the war years perhaps played their part less directly in shortening the lives of several of his older friends. Macdonell's death in 1916 was followed by that of Lord Parker of Waddington in 1918 and of Goring in 1919. Robert Parker had been Pearson's closest College friend, they had shared chambers in the Temple and joined in the founding of the men and women's club to which I have referred. Their careers had followed widely different paths, but the lawyer and judge had retained a warm interest in his friend's more adventurous course and had been one of the five original guarantors of *Biometrika*. Of Charles Goring I have already spoken; he was a man for whom more than any one else Pearson would have liked to find a post in his laboratories. Those four years of physical and emotional stress must have had some effect, too, on Pearson's own vitality. He had still his old power to inspire those who worked under him and the strength to push forward eagerly with his own ideas, but perhaps it was harder now for him to keep abreast of the ideas of others, or to step out from his study to defend the causes in which he believed.

In a more material sense the war period had indeed dealt a hard blow to the prospects of the Galton Laboratory. Funds which would have proved adequate to equip the new building in 1914 were now quite insufficient to meet a 300 per cent. rise in costs; the pre-war salaries of his staff were on a scale that in changed conditions ceased to provide a living wage. Nevertheless, slowly, with help from many generous friends, with the aid of a public appeal, with added contributions from University sources, the situation was mastered. The London County Council agreed to provide funds necessary for the salary of a medical officer and his assistant, a post created, alas, too late to be offered to Goring. The Medical Research Council gave an annual grant to Dr Julia Bell to enable her to carry on investigations into the inheritance of disease. The Drapers' Company continued their support. And so in October 1919 Pearson and his staff were able to occupy the first floor of the Bartlett building. By the end of that Session the whole territory

had been occupied, although parts of it were still very sparsely furnished, and the total staff of the twin Biometric and Eugenics Laboratories had risen to ten persons, a figure at which roughly it remained during the following years.

At this period Pearson was giving much time to the completion of a long-delayed monograph, "A Study of the Long Bones of the English Skeleton" **(50)**, written jointly with Julia Bell, and to the allied papers with Adelaide G. Davin, "On the Sesamoids of the Knee-Joint" **(51)**, **(52)**. Pearson's original object had been to show that a precise technique of measurement, leading to biometric methods of analysis and comparison, could be developed for the long bones of the skeleton as for the skull. While circumstances had limited the scope of the research in the sense that it had finally been concerned almost entirely with the femur, it had been extended before completion in another direction. Not only was comparison made between the English femur and that of other races, but ultimately, after following back the development from recent man to palæolithic times, Pearson was drawn into a study of the femora of the Primates. Thus the Chimpanzee, the Gorilla, the Orang, the Old and New World Monkeys, the Lemuroids and Tarsius were considered. This use of the intensive study of a single bone to throw light on the development of man formed the subject of Pearson's lecture of May 14th, 1920, to the Royal Institution, entitled "Side Lights on the Evolution of Man" **(96)**.

In working at the femur many data had been collected regarding the sesamoids of the knee-joint; with the evolutionary problem in mind and a trained biologist at hand in Miss Davin, Pearson extended his investigation to the case of many mammals, birds and reptiles and produced a paper beautifully illustrated by Miss Davin herself and by Miss Ida McLearn, who for a number of years was to carry on the Laboratory tradition of good draftsmanship. The object of the paper, it was stated, was "to suggest problems to those better equipped for studying them than the present authors, rather than to present solutions." Might not the least important of little bones, which anatomists following Galen had passed by as of no interest, contribute their own special share to the general assault on the enigma of evolution?

With interest keenly roused in such work as this, the war-time memories were receding. On June 4th, 1920, the new building was formally opened by the Minister of Health, and the Department of Applied Statistics set out afresh on a new venture.

1920–1933

When we mind labour, then, then only, we're too old.
Robert Browning.

The post-war years were not favourable to the spread of Galton's eugenic creed. Too much idealism had been poured out vainly in the battlefields of France. Men were tired by war, disillusioned with peace, restless, too much concerned with troubles of the present to be ready to plan a course aiming at distant horizons.

The war gap had made a break there, without doubt; it is likely that Pearson him self, with the world's reckless spending of its best in mind, was too conscious of the irony involved to return with any eagerness to the task of urging man on to the improvement of his breed. But if interest in Eugenics was for the moment set aside, there was a growing call for the use of statistical method. The Medical Research Council invited Pearson to be chairman of their Advisory Committee on Statistics*; it was a post that he felt unable to accept because of the large amount of work that must be done in his own Department, but the recognition which it implied of his many years of effort was heartening. In the Sessions 1920–22 the lecture courses on the theory of statistics at University College were very well attended. The audience was drawn from many directions; it included several members of the staff of the Industrial Fatigue Research Board, an Indian Civil Servant, several Cambridge mathematical graduates, a number of Americans including two professors on sabbatical leave, and finally five undergraduate students working for the new Honours Degree in Statistics.

The institution of this Degree by the University in 1915, as a result of Pearson's initiative two years earlier, meant that the subject of mathematical statistics had now an official standing in the University. There followed also some change in the programme of the Department; lecture courses which in the past had been arranged somewhat informally to suit postgraduate students had now to be planned in accordance with a definite syllabus. There were main first and second year courses in statistical theory given by Pearson himself, each consisting of two one-hour lectures a week with several hours of practical work and, in addition, auxiliary courses taken by his assistants on probability, on interpolation and quadrature and on periodogram analysis.

Pearson's feelings regarding the introduction of an undergraduate element were mixed. He would have welcomed a steady supply of three or four good students a year; some of these might remain as research workers, the rest he felt could be sure of finding posts outside. He regretted the refusal of the University to grant an Intermediate Examination in Statistics, which prevented students being brought into touch with an interesting subject at an early stage in their careers. Nevertheless, when some years later a very young generation full of youthful spirit romped down his polished stairs or trundled each other round the class-room in a small hand-cart meant for carrying books, he felt justly indignant. He often had cause, too, to grudge the time which his staff must give in teaching students how to handle their practical work and, in particular, in finding out their mistakes. He was anxious that the Department should not be swamped by undergraduates; it must remain a "research," not a "teaching," department.

The written record of Pearson's lectures remains only in the note-books of his students; his was not a style that would have led easily to a text-book, though

* This was a second public recognition by the medical profession of the position which Pearson occupied, although outside its ranks; in December 1919 he had been elected an honorary fellow of the Royal Society of Medicine.

many tried to persuade him to write one. As his old engineering student, whom I have quoted, said*: "K. P.'s methods were secondary to his personality—that was the key to his success, in the keenness and interest his students took in all his classes." I believe, however, that it is worth putting on record (in Appendix IV) a summary of the subjects covered in his two lecture courses on the theory of statistics given in the Session 1921–22, for which my own lecture notes taken at the time are available. In discussing these lectures, their manner of delivery and the subject-matter they contained, I have made free use of the impressions with which other students of that and somewhat later periods have kindly supplied me†.

The new student, accustomed to lectures elsewhere, was perhaps surprised that we started with reference to no text-books; but as the course progressed we were referred to volumes of *Biometrika*, to the *Tables for Statisticians* and to papers in the Royal Society publications. We heard much of Galton, something of Yule, of Edgeworth and of Sheppard, but of other writers little, and that sometimes with a warning. We were told of the sins of many people: of the compilers of Government Statistics; of the writers on the Theory of Errors, who had illustrated the Normal Law on data which in fact gave a fit of 1000 to 1 against; of the astronomers, the psychologists and the chemists who used hopelessly small samples; of the anthropologists who would not recognise the value of biometric methodology. But if Pearson spoke critically as one having authority, he would humorously admit his own errors and he would give us lavishly, what after all we really wanted, his own views on things; he inspired us by showing his creative mind at work, throwing out at the same time frequent hints of problems that needed still to be solved. Several of these were, in fact, taken up by his listeners.

He usually took some pains after dinner on the evening before to prepare his morning lecture; he would come into the class-room with these notes, put them on the table and then proceed to lecture for an hour without ever, save on very rare occasions, referring to them; even then it was probably only to see what was the next subject with which he proposed to deal. During the hour, the board would be filled several times over with algebra and diagrams, all of which it was possible for a keen student to take down in notes in a form which was afterwards readable and lucid. He would plunge sometimes with seeming delight into a piece of long and heavy algebra, retaining all the interest of his class while doing so and imparting a certain spirit of excitement and wonder as the complexities were untied into a neat solution. He had the lecturer's gift of appearing to discover a result for the first time himself. If, as happened now and then, he saw that his equations were coming out wrong, he would pause, regarding the board reflectively, with the remark "Let's be cautious, let's be cautious," and he showed evident satisfaction when the correct result was obtained.

* Part I of this memoir, p. 208.

† E.g. Mr C. H. L. Brown, Dr J. O. Irwin, Mr Frank Sandon, Miss Brenda Stoessiger (now Mrs Clapham).

He was sensitive to the response of his audience. Many will remember how he watched anxiously to see if they had followed some special point that he had made, and the pleasure which he showed on finding that they had grasped it. Speaking of one of his foreign professors he would say, "I like old ——, he always beams back at me when I make a nice point." While he was sometimes temporarily annoyed at being interrupted in the middle of writing a formula, he welcomed intelligent questions and blamed his class for lack of attention if they allowed an algebraic slip of his to remain unnoticed for several lines.

Until the last two or three years he kept in touch with every student in the Department, making it a practice to go round the laboratories at least once a day and say a few words to each about their class work or research. Such visits would extend sometimes to half an hour or an hour, K.P. sitting down with pencil or pen to illustrate his remarks or to look for a puzzling error. He was sympathetic to the difficulties and criticisms of his younger research workers and students; sometimes he would spend much time in a subsequent lecture discussing more fully a question on which they had expressed doubt. As I have mentioned in an earlier section, it was, however, sometimes more difficult to thrash out a difference of opinion with him at a later stage of one's career.

It will be interesting, I think, to give some attention to the subject-matter of the first and second year lecture courses with which these recollections are associated, because we shall find here evidence of the stage to which by 1921 the mathematical theory of statistics had been built up in England, largely through Pearson's own efforts. Turning to the summary in Appendix IV, it will be seen that the course began with an outline of the conception of correlation. I find on page 1 of my Notes the following statement, which was probably taken down fairly closely from Pearson's words:

"The purpose of the mathematical theory of statistics is to deal with the relationship between 2 or more variable quantities, without assuming that one is a single-valued mathematical function of the rest. The statistician does not think that a certain x will produce a single-valued y; not a causative relation but a correlation. The relationship between x and y will be somewhere within a zone and we have to work out the probability that the point (x, y) will lie in different parts of that zone. The physicist is limited and shrinks the zone into a line. Our treatment will fit all the vagueness of biology, sociology, etc. A very wide science."

It was this idea of correlation, first drawn from Galton's *Natural Inheritance*, which stood for Pearson as the fundamental illuminating conception of the statistical calculus. Just as another man might have taken as a leading *motif* the application of the abstract theory of probability to the solution of problems in the perceptual world, or the assignment of total variation into parts attributable to various causes, so it always seems to me Pearson took this concept of correlation. He referred to it often in his written work; thus we find in the new fifth chapter

on "Contingency and Correlation—the Insufficiency of Causation" added to the third edition of *The Grammar of Science* ((12) p 170) the following paragraph:

"As a method of predicting the experience *likely* in the future from the experience of the past, the summary of the past expressed by function or under the category of causation has done immense service. But it is incomplete in itself, for it gives no measure of the variation in experience, and it has trammelled the human mind, because it has led to a conceptual limit dominating actual experience. We have tried to subsume all things under a perfectly inelastic category of cause and effect. It has led to our disregarding the fundamental truth that nothing in the universe repeats itself; we cannot classify by sameness, but only by likeness. Resemblance connotes variation, and variation marks limited not absolute contingency. How often, when a new phenomenon has been observed, do we hear the question asked: What is the cause of it? A question which it may be absolutely impossible to answer, whereas the question: To what degree are other phenomena associated with it? may admit of easy solution, and result in invaluable knowledge."

Those who attended his lectures will remember that characteristic correlation belt which he drew on the blackboard and the evident pleasure with which he filled its broad zone with dots, each tap of his chalk hammering, as it were, a fresh nail in the coffin of causation.

Returning to the details of the course of 1921, it will be seen that he passed from correlation and contingency through linear regression to non-linear regression, and then turned to consider the form of the variation within each array about the regression line. There was a certain grandeur in this planning which led to the consideration of variation as a feature of correlation; it was something which captured the imagination in a way which the more usual scheme of starting with the description of a single variable and proceeding to study the relationship between two and then three etc. does not The early introduction of orthogonal polynomials did not, however, make the going easy for undergraduate students.

About sixteen lectures were devoted to Pearson's system of frequency curves. It was a part of the subject with which he particularly enjoyed dealing and he was able to convey this interest to his audience. As one of his post-war students writes:

"If I had to single out any part of the course which has remained most of all in my memory, it would be the wonderful *unity* of the system of frequency curves and the logical development of the subject, each curve being derived from the fundamental differential equation according to the assumptions made regarding the values of the constants in that equation....K.P.'s own enjoyment at seeing a set of data accurately represented by a frequency curve was a pleasure to watch."

In laying considerable emphasis on fitting frequency curves, Pearson was concerned not only to provide his students with a useful tool, but to give them a training in method and accuracy; the work gave an opportunity for practice in the use of tables, the application of interpolation and quadrature formulae and even

of draftsmanship, since, at any rate in the earlier days, students plotted their curves and were shown the use of a spline and a planimeter.

There was one aspect of Pearson's approach to his frequency curves on which I have never been quite clear. The fundamental differential equation could be derived by considering the ratio of the slope to ordinate in a hypergeometric series, or, as a particular case of this, in the binomial series. He had seemed, at any rate at one time, to feel that there was some physical link underlying this relation which gave a special significance to the resulting curves. In later years he would still refer to this aspect of the matter, but he laid more stress on their proved usefulness as graduation formulae. Their value in another field, that of sampling theory, was first shown in "Student's" paper of 1908* in which a Type III curve with the correct moments was used to represent the unknown true distribution of the variance in samples from a normal population. The system has been used increasingly in recent years for similar purposes. Finally R. A. Fisher's work has shown how the mathematical forms of the Type I, II, III, VI and VII curves give the exact sampling distribution of various criteria used in the analysis of variance.

When dealing with the problem of graduation there are sometimes undoubted difficulties in fitting the curves and there may be genuine differences of opinion as to the most serviceable method to employ in this process; but a perversity which denies both the practical utility and theoretical interest of the system can only result in a serious curtailment of useful tools in the statistician's workshop.

After dealing with frequency curves, Pearson approached the problem of statistical inference, "the fundamental problem of statistics being," he said, "to predict from the past what will happen in the future." He spoke much as he had written in the latter half of chapter IV of *The Grammar of Science* and in the Gresham Lectures, and he added the theoretical work on the extension of Bayes' Theorem that he had recently published (**97**). This paper, as Pearson afterwards recognised (**98**), contained a curious oversight, two distinctly different frequency functions being assumed identical, and an unduly simple result obtained in consequence. Perhaps it was due to a temporary lack of clearness in thought, a fault to which, I suppose, all of us succumb at times! In any case in this particular instance the error acted as a stimulant on various members of his class; we discussed and argued and experimented with it, and from the new ideas which were set in train we were the gainers.

The latter part of the first-year course was concerned with the study of probable or standard errors. The line of approach to this subject had been determined by the types of problem with which biometry had been faced in the past; it had been necessary to compare populations from which large samples could generally be obtained, or to determine the effect of selection of one character upon the variability or intercorrelation of other characters. With large samples, first approximations to standard errors were adequate and these could often be obtained by the very

* *Biometrika*, VI, pp. 1—25.

serviceable method of expansion of the variable measure in terms of "statistical differentials." As far as possible, Pearson avoided any initial assumption about the form of the observed variates. Thus the standard error of a squared standard deviation was first expressed in terms of the β_2 of the sampled population; as a special case, β_2 was set equal to 3.

A division was set between large-sample and small-sample theory, where later work has tended to emphasise continuity. For example, I find the following sentence in the 42nd lecture: "It is at this point that the whole theory of small samples diverges from our subject, and what may be said in the near future as to the probable errors of frequency constants will be based upon the assumption that our original population is large and that the sample is not small." It is interesting to note, too, that while Fisher's derivation in n-dimensioned space of the distribution of the squared standard deviation, s^2, was given, the result was used to examine how quickly the approximate formula for the standard error of s might be employed and how soon the sampling distribution might be taken as effectively normal, rather than to consider applications in connection with small samples. Again "Student's" distribution of z was not discussed in these 1921–22 lectures*, because Pearson felt no interest in work based on small numbers. No doubt if he had been faced earlier in life with the type of problem that *must* be tackled in an experimental brewery or in an agricultural research station his views on small samples would have been very different.

The second-year course dealt with: (1) Partial and multiple correlation and the application of this theory in a number of directions, including problems of ancestral heredity. (2) The χ^2 test of goodness of fit, which was derived from the multiple correlation distribution. (3) The simple principles of the Mendelian theory and the resulting correlations of heredity to be expected in a population mating at random. (4) The development of probable error theory in the case of a bivariate distribution. (5) Methods of estimating correlation from variables classed in broad qualitative categories. (6) A number of miscellaneous problems such as the variate difference method, Galton's individual difference problem, etc.

Throughout both courses papers containing illustrative examples to be worked out were issued in connection with each piece of theory. Possibly these examples involved the students in too much heavy calculating labour, but the training was designed to make quick and reliable computers, unafraid of any task and familiar with the use of a wide variety of methods.

Looking back now at these courses after fifteen years, one is struck inevitably by the big developments in statistical theory that have intervened. Yet I believe that those of us who received Pearson's training, and have since owed a great deal also to the stimulus of R. A. Fisher's new ideas, are aware of an essential continuity which runs through the central line of English mathematical statistics. As I have suggested above, does not the greatest change lie perhaps in the interpretation of

* This remark does not apply to Pearson's lectures in later years.

the meaning of probability? In applying the mathematical theory of probability, involving abstract concepts, to the interpretation of observational data, there can be no unique solution, no right or wrong. Within the field of mathematical theory there can be errors, but how we use this conceptual model to help in determining our decisions in the world of experience must remain to some extent a question of individual choice. Probably it was a failure both on the part of Pearson and of his critics to recognise this distinction which led to much controversy on the use of χ^2 and other matters.

A single observed event, whether the frequency of individuals in a class, the value of a mean $\bar{x}$ or of a measure of discrepancy such as χ^2, cannot provide a measure of probability until we have defined the set of objects, or what has been termed the fundamental probability set, to which it belongs. And here there may be considerable variety of opinion as to the most appropriate set. It is worth while trying to throw some light on this point by considering a simple concrete example, because we are, I think, concerned with two distinct views, one of which seems now on the whole to be regarded as more useful than the other.

Consider the case of a statistical test to determine whether the population mean of a variable x, known to be normally distributed, can have a specified value ξ, when a random sample of n individuals gives a mean $\bar{x}$ and standard deviation s. Pearson, following what may be described as the classical tradition, considered the fundamental criterion to be

$$\frac{(\bar{x}-\xi)\sqrt{n}}{\sigma} \tag{i}$$

where σ was the population standard deviation. If σ were unknown (remembering that in large samples s would be close to σ), he considered that he could form his opinion as to the likelihood that the population mean was ξ by obtaining from the tables of the normal probability integral an answer to the following hypothetical question: How often could as large or a larger value of

$$\frac{(\bar{x}-\xi)\sqrt{n}}{s} \tag{ii}$$

be expected to occur in random sampling from a population with mean $=\xi$ and standard deviation $=s$? It was the set of possible samples from this population which formed the fundamental probability set.

The new approach was to regard the value of the ratio (ii), calculated from the sample, as belonging to a different probability set, namely the set of values that would be generated if we took repeated random samples from normal populations with mean at ξ and the same or different σ's, and inserted each time the observed sample s into the ratio. The old approach led to the tables of the normal distribution, the new to those of "Student."

It will be seen that in both cases the set of objects to which the single observed ratio is referred is really conceptual, not experiential; yet many of us feel now intuitively that the second set is the more appropriate, largely perhaps because of

the direct correspondence between the probability measure found from the tables and the risk of error which will follow in the statistician's long-run experience, if he calculates ratio (ii) and rejects the hypothesis he is testing at a given probability level. I believe a closely similar change of approach can be traced in the use of many others of the tests with which Pearson originally provided us, e.g. that of χ^2. It is a change which has seemed to a new generation to throw a wealth of illumination on to many problems, to bring results hitherto disconnected into a new perspective. We felt that we had fallen on one of those simple and clarifying laws of thought which Pearson had described for us in his *Grammar of Science*. It is quite certain that he did not himself feel that simplification; nor was he alone in this. He had trained his mind to regard probability in a different way and at the age of 65 or 70 the new ideas, often somewhat obscurely presented, were not easily grasped or if grasped did not appear very profitable. We cannot say he was wrong nor that his opponents were right; it well may happen that twenty or thirty years hence our own views of to-day will have succumbed to a fresh outlook or even to something like the old outlook. But where perhaps he was at fault was in a failure to recognise that a younger generation was as genuinely chasing along new lines of thought as he had been himself in the '90's. The old watch-dog should have let them pass!

I have spent some time on this discussion of statistical theory because those who have studied the pages of *Biometrika* since the war will realise the important part that the development of theory still held in Pearson's thoughts. He was far from satisfied with the position he had reached. He saw, he thought, how further advances could be made, but knew that his mathematics were not now adequate for the task. "I wish I were young again, I wish I were young," a friend remembers him saying, standing one day on his hearth rug, smoking an after lunch cigarette and looking past her along the pathway he could see, with that characteristic upward tilt of the head. Up to that last suggestive paper of December 1933 (**99**), on what he termed the P_{λ_n} test, he was continually coming back to questions of theory. These papers are full of fresh ideas, but to study them with profit the reader may need to have some understanding and sympathy with an approach which differs from his own.

If we pass in review the many enterprises freshly entered into and satisfactorily completed under Pearson's direction between the official opening of the new Galton Laboratory in 1920 and his retirement in 1933, it is difficult to realise that this period covered the interval between his 64th and 77th years. In this connection he would certainly himself have reminded us of Galton, who published his *Natural Inheritance* at the age of 67 and subsequently propounded the science of eugenics. But Galton was to a large extent writing and thinking in the peace of his own study, while Pearson had to cope from day to day with the active direction of a research and teaching department. I shall not attempt any detailed description of these thirteen years, partly because they are still too near in time and partly because

it is not easy for one who was a member of the staff to place in proper perspective the department's many activities*. It must suffice to give some account of the lines of work most clearly stamped with Pearson's own individuality; among these the following are of particular interest: (i) *The Tracts for Computers* and Statistical Tables, (ii) Dog Breeding and the Animal House, (iii) the Anthropometric Laboratory, (iv) Craniometry, (v) the Museum, (vi) *The Annals of Eugenics.*

(i) *The Tracts for Computers and Statistical Tables.* The origin and objective of the series of *Tracts for Computers* is given in the Editor's prefatory note to No. I **(100)** as follows:

"During the course of the past five years the Department of Applied Statistics in the University of London has carried out a great deal of computing work of one kind or another bearing on special war problems of a physical character. Its members have been struck by the absence of any simple text-book for the use of computers and still more by the absence of obviously necessary auxiliary tables. The present series of *Tracts for Computers* will endeavour to fill this gap as far as it lies in our power. It will not concern itself with the higher mathematical theory, but solely with the practical difficulties of the computer, or rather such difficulties as we have met with in our own experience. The first tract will be followed not only by others containing recently computed tables or by the re-publication of old tables at present very inaccessible, but by tracts dealing with interpolation, quadrature, mechanical integration, calculating machines, tabling machines, and bibliographies of memoirs and of tables having special value to the practical computer."

The series ultimately contained such widely different works as Henderson's Descriptive Catalogue of Mathematical Tables (only partly completed), Tippett's Random Sampling Numbers and A. J. Thompson's monumental 20-figure Logarithms. I think that Pearson felt that on each generation of mathematical workers there lay a duty to contribute a share in the gradual building up of a corpus of accurate mathematical tables. In a long life he played his part nobly with his own Brunsviga; it was his custom, too, to put each research worker on to a piece of table computing at some point of his training. The successful achievement of an exacting though dull piece of work was in his view a good test of character. *The Tables of the Incomplete Gamma-Function* **(101)** 1922, *The Tables for Statisticians and Biometricians*, Part II **(102)** 1931, and *The Tables of the Incomplete Beta-Function* **(103)** 1933, were the most important results of many years of long co-operative labour.

(ii) *Dog Breeding and the Animal House.* When at last, in 1922, the Department gained possession of an Animal House, it was too late for Pearson himself to plan any ambitious new biometric research. Indeed the building itself, an old cabman's house overrun with wild mice and shadowed by other buildings in what had once been a mews at the back of the College, was but a makeshift affair until funds were supplied for its rebuilding in 1930. It did, however, make it possible to carry on under easier conditions the dog-breeding experiment. The animals which before had

* In the five years 1925–29 alone, 130 original papers appeared from the Department of Applied Statistics; of these, 36 bear Pearson's name as author or part author, but he took a hand in the writing up for press of many more.

been placed out with various persons in kennels often far from London, could now be kept together under observation. It added also a pleasing variety to our work; we could go over after lunch and hold a puppy while the Professor measured its nose and head; or in more strenuous fashion, when the Animal Boy was on holiday, scrub pens, cut up meat and exercise the dogs in turn in the yard. Black, white, red, yellow and piebald, they were a goodly collection of young creatures, although in later years, perhaps from too little sun and too much inbreeding, they became less robust and harder to breed from.

Almost from the very beginning of the experiment in 1908 Pearson had kept some of the dogs himself at home and many litters were born and weaned at Hampstead, at holiday houses and at Coldharbour. Now this domestic burden could be ended, but two or more of the older animals were always there as old friends about the house. How many of us can remember one or other of that long succession! Tong, of the foundation stock, Ling, her son, who as a puppy had lived with Galton and who was afraid of no man or dog however large, Choo, Hans and Grethel, Donach Ruadh, Meg and her son Ben, Topsy, Kar and Eld, Shagpat and Gemima. To their master, their characters were as interesting and varied as their coats; and if at times they gave cause for anxiety when their barking interrupted work, when they strayed into the road, when they ran on the garden beds and, if science demanded, when they failed to propagate their kind, nevertheless they were a very real source of pleasure.

(iii) *The Anthropometric Laboratory.* A generous gift from a member of the staff made it possible to start the equipment of this Laboratory in 1921. Its first purpose was to obtain records from tests on the College students. The sight, the hearing, the judgment, the mental agility, the strength, the physical measurements in a great variety of forms were taken and a wide record of faculties preserved. It was hoped also to associate these results with University achievements. The work was in charge of Dr Percy Stocks, but when in full session the help of all the staff was needed, each taking charge of two or three tests. Until trouble with his eyesight prevented him, Pearson was responsible for the head measurements. After the novelty had worn off we found it a heavy job and in time there was difficulty in obtaining an adequate supply of students. Perhaps we had too many tests, but Pearson felt that one could never be sure of what was most important in a research until a full record of data was available for analysis; better to discard then, than to fail because of inadequate material. I have often admired the patience with which Dr Stocks collected his somewhat reluctant band of workers and spent many lunch-time hours himself bearing the brunt of the labour.

(iv) *Craniometry.* With Pearson's contribution to craniometry I am not fully competent to deal. I have given above, however, some indication of what I believe were the objectives of the long-time research that he planned: the accurate description of racial characters within each group and, as more material became available, a critical study of the relationships between groups. The Department became a store-

house for more and more series of skulls, some supplied from Egypt by Flinders Petrie, others obtained from the London area when new building opened out an old burial ground or plague pit. The collection of about 7000 skulls which Pearson left at his death is something of unusual value.

In the years after the war he was to find an able and sympathetic collaborator in Dr G. M. Morant, who played an essential part in the work of the craniometric laboratory, taking most of the responsibility of training the research workers off Pearson's shoulders and cheerfully bearing much of the heavy labour of getting their papers in order for the press. Morant was also in later days responsible for photography; money for the equipment of the photographic rooms was never obtained, but he and others made good use of an old full-plate camera, a tried friend of the Biometric Laboratory for more than thirty years.

Pearson's own contributions, among which may be mentioned a paper "On the Biometric Constants of the Human Skull" (**104**), papers on the Coefficient of Racial Likeness (**105**), (**106**), an Oxford lecture on a new Cranial Coordinatograph (**107**) and a study of the individual bones of the skull (**108**), are characterised by a rare suggestiveness and breadth of view. Of some other original uses to which he turned his interest in physical anthropology I shall speak below; here it may be of interest to note that the period of this section of this memoir started with his presidency of Section H (Anthropology) of the British Association in August 1920 and ended with a lecture, his last public lecture, to the Oxford University Anthropological Society in May 1933. Further, in 1932 he was awarded, as the first foreign recipient, the Rudolf Virchow medal of the Berlin Anthropological Society.

(v) *The Museum.* In the original plan of the Francis Galton Laboratory, there was a door into the street through which on certain days the public would have been admitted to a museum bearing on the subject of Eugenics and thence enticed into an anthropometric laboratory. The authorities decided that an independent entrance to the College could not be permitted and so this plan had to be abandoned; nor, indeed, after the war were funds available to provide the staff needed to deal with the public in this way. There was, however, space provided for a large museum and this was gradually equipped with cases, through the help of many private benefactors. The range of exhibits was typical of the wide interests of the Director. In deep cases down one wall were statistical models, some dating back to his first attempt in the Gresham College lectures on Geometry to make clear the meaning of probability theory to a popular audience; others of more recent date dealt with advanced statistical theory. In revolving show cases was a collection of photographs, drawings and pedigrees, illustrating the inheritance of good and bad qualities in man; pedigrees of ability, as in the Bach, the Maclaurin and the Bernoulli families; pedigrees of defect, whether of insanity, digital deformities, congenital cataract or haemophilia; photographs of albinos, of dwarfs and of lobster-claw deformity. There was also a special section dealing with the early history of man; artefacts, prehistoric burials, casts of bones and of some of the more famous prehistoric skulls; also copies

of the reconstructions by Mascré of Brussels of certain types of early man. Other cases contained type skeletons of the Pekinese and Pomeranian dogs and of the various degrees of their crossing. There were skins of these dogs, too, and of the hares discussed in the *Albinism** whose coat colour changes with the season; paintings of albinotic eyes; colour scales and examples of various measuring instruments.

There can have been few museums that bore a closer, a more striking witness, to the activities and interests of a single mind. The sympathetic visitor who was fortunate enough to be taken round by Pearson himself did not easily forget the occasion.

(vi) *The Annals of Eugenics.* Several series of popular lectures on the work of the Biometric and Eugenics Laboratories were given after the war; for example, Pearson gave a course of 10 lectures in the spring of 1921 and other courses in which the members of the staff took part were given in 1922, 1924 and 1929. But popular interest was far less keen than before the war and it seemed that there was little prospect of success at that moment for Galton's plan of "extending the knowledge of eugenics by occasional public lectures." Pearson's last great effort towards the establishment of eugenics as a science was the founding in 1925 of a new journal, *The Annals of Eugenics.*

"The time seems fully ripe for the issue of a journal," runs the editorial Foreword, "which shall devote its pages wholly to the scientific treatment of racial problems in man. Several journals allot some of their space to original memoirs dealing with eugenics and the general problems of race hygiene. Others of a minor character spend their main energies in popular articles, book-reviews and accounts of matter published elsewhere. Our journal will differ from existing journals in that bibliographical matter will be reduced to a minimum, that no other topics than the problems of race in man will be dealt with, and that the papers published will be the work of trained scientists rather than of propagandists and dilettanti. Naturally a journal issued by the Galton Laboratory will be sympathetic to the methods of its founder summed up in the title of his Herbert Spencer Lecture 'Probability the Foundation of Eugenics.' But this does not signify that contributions dealing with heredity in man from any scientific standpoint will not be acceptable Nevertheless the study of man is essentially a study of mass-movements and mass-changes. Selection can hardly take place in man except by selection of somatic characters, and the results of such selection can only be effective as an evolution, according to the extent to which somatic and germinal characters are correlated. The existence of such a correlation is an undoubted fact, whatever theory we may choose for its expression...."

"It may be argued that there is a science of Eugenics which is not our Eugenics, and if one might place faith in the multitude of text-books, which have adopted the name, the argument would be complete. Most of these text-books, however, have merely taken the name and nothing else from the founder; they mix a little biology with a trifle of genetics, and water the whole down with much tea-table talk on the impracticability of fundamentally improving the race of man. In such manner no great science was ever

* (49) Part II, pp. 421—442.

built up; it must have definite methods of attack—be in short a concise discipline—realise its problems and grasp how their solution can be approached. It was such a discipline that Galton foreshadowed when he claimed that probability was the foundation of Eugenics...."

"It is the aim of our journal to aid, as far as lies in its power, the oncoming of the day when we can claim that the groundwork of our science has been securely laid, and both the student's text-book and practical eugenics—eugenics applied to national problems—will then be feasible. Let us bear in mind the words of Galton written almost in the last years of his life, words not of despair, but of wise caution: 'When the desired fulness of information shall have been acquired, then and not till then, will be the fit moment to proclaim a "Jehad" or Holy War against customs and prejudices that impair the physical and moral qualities of our race.' This has been the spirit in which the Laboratory he founded has been conducted, and that will be essentially our guide in the control of this journal."

It was inevitable that at first the great majority of the contributions to *The Annals* came from inside the Department of Applied Statistics; memoirs of a kind which before had been issued in the Drapers' Company *Studies in National Deterioration* and the *Eugenics Laboratory Memoirs* were now published in the new periodical. But Pearson looked towards the future, to a time beyond the days of his own editorship, when there should be a steady expansion in the field of workers, "when every university of standing will have its professor and laboratory of Eugenics."

Of this journal he edited, with Miss E. M. Elderton's assistance, five volumes. His own chief contribution in its pages to the scientific treatment of racial problems was the long piece of statistical research on "The Problem of Alien Immigration into Great Britain, illustrated by an examination of Russian and Polish Jewish Children." This work, in which he was assisted by Miss Margaret Moul, appeared in five parts, and was not even then completed (82). It was based on the data collected before the war, regarding children at the Jews' Free School to which I have already referred. Three of Pearson's last public lectures, "On...the Relationship of Mind and Body" (109), "On a New Theory of Progressive Evolution" (110) and "On the Inheritance of Mental Disease" (111) were also issued in this journal.

Characteristically, he put the adequate presentation of his subject before the convenience of his readers. The unusually large-sized page of those first five volumes was planned to prevent any cramping of the space for tables, diagrams and skull photographs.

At the heading of an earlier section of this memoir I have set some words of Michael Angelo's, one of the mottoes on the walls of Pearson's room, which imply that the true pursuit of science can leave no man free for distractions. Nevertheless, Pearson's life was full of the practice of what may well be termed hobbies: as a professor of Applied Mathematics, biometry and astronomy had been his hobbies; so, when he became professor of Eugenics, had been research into the ancestry of Darwin and of Galton. And now, in his later years, his mind played at large in

many fields. In the speeches that he made and the toast cards with which he surprised us at the annual Galton Dinners; in the portraits and cartoons he collected for the walls of the Department; in the Museum, the models, the instruments, the busts; in his lectures on the History of Statistics; in his papers on the skulls of Robert the Bruce, George Buchanan, Henry Stewart Lord Darnley and Oliver Cromwell; in all these things we were aware of the richness of his mind.

The first of the Galton Dinners was held in 1920, the fourteenth and last in 1933; the date was as nearly as possible January 17th, the anniversary of Galton's death. They were evenings when the past and present workers and students in the laboratories collected to see and talk to their Professor and to each other, and to find themselves again in that atmosphere, undefinable but very real, which had first grown up round the little one-room biometric laboratory of the '90's. Besides "ourselves", among whom must be included a number of friends and benefactors of the Laboratories, one or two guests of distinction were welcomed on each occasion. All those who attended will remember the characteristic features of those dinners; the assembly in the Museum, K. P. greeting us by the door, Miss Elderton hurrying round, with smiles for us all, to introduce neighbours at the dinner table who were unacquainted with each other; the dinner itself in the College Refectory, the file back, two by two, along an underground passage to our own building; the roll-call in the Museum to ensure that we went upstairs in the right order to squeeze into our seats for dessert and toasts in the Common Room, or later, when numbers became too large, in one of the student laboratories.

There were at first five Toasts, afterwards increased to six: (1) In pious memory of Francis Galton, (2) In remembrance of all Benefactors, (3) In memory of the Biometric Dead, (4) The Guests, (5) The Postgraduate and Student Workers, (6) The prosperity of the Laboratory. In Pearson's speeches he would tell us of Galton, of the founders of *Biometrika*, of little incidents from his long memory of the history of biometry; of losses we had suffered by death since the last meeting; of the work carried out in the Department during the past year. There was a vein of sadness running through much of what he said, inevitable perhaps in one who looks back to recall the contemporaries or even younger friends whom he has lost. But he could deal with the present in a lighter manner. Below are some extracts from his Toasts of 1922, a year selected at random from the series.

In pious memory of our Founder.

"The present year is one of great importance in our annals. On February 16th it will be 100 years since the birth of Francis Galton. A century which deserves celebration at least equally with that of his cousin, Charles Darwin. For while the ideas that sprung from Darwin led to a reconstruction of all the biological sciences, those that sprung from Galton's inspiration are leading to a revolution in scientific logic; they must ultimately produce a renascence in every science in which statistics plays a part, and that we may truly say involves every branch of modern knowledge. But I do not desire to dwell on that side of Francis Galton to-night. I would urge rather another phase of his work, his power of exciting the affection of the most divergent men and women...."

In pious memory of Sir Herbert Bartlett and other benefactors.

"It is fitting that in this year when the donor of our new buildings has died, we should single him out for special reference and honour....

"Summing up what benefactors have done for us, we can say we have now our building and practically nearly the whole of its fittings as originally designed. We have still much to do by way of equipment. I remember the day when two machines had to serve the staff of four men and one woman, and she, our dear old Dr Lee, whom illness has kept from us to-night, went out in indignation and bought herself a machine. I regret, I cannot even say out of her stipend, for she served the Laboratory—and kept us all in proper discipline—for nearly 15 years without a penny of payment. She gave us at present rates of pay between £2000 and £3000, and possibly over 1000 persons are now using her tables and do not recognise her as a benefactor, who has saved them more than they will or can ever repay her.

"Well, we do not have two machines to five computers now-a-days, but the equipment, as we all know, is still defective ..."

In pious memory of dead biometricians.*

"I will not say more this year than again couple this toast with the names of Galton, Weldon, Macdonell and Goring...."

To our guests.

"...Dr Katherine Watson..., Mr Hope-Pinker..., Dr Whitehead..., Mrs Hume Pinsent .

"And lastly I come to Sir Gregory, our Provost. I have already referred to him to-night. It is only human nature to abuse the head of the executive, if there are no towels or the windows don't get cleaned. Every failure in the government machine is of course Mr Lloyd George's fault. But when the foreigner begins to criticise our prime minister, then we are apt to see his virtues, and reply hotly that from Lenin to Poincaré we don't see a better. And looking all round the colleges and universities of the country, I don't see a better head than ours Although having known him since I was in knickerbockers and he was in—well, in the perambulator—I may at times give way to the instinct to pummel him if he exercises his legitimate right to disagree with me.

* The following unfinished lines, written on the theme of this toast by Mrs Pearson, were found in a book by her bedside after her death in 1928:

(1)

In a quiet sheltered land, a land of Silence,
Where the spacious courts and halls of Memory spread,
Waiting till we turn and seek their counsel,
Rest our Biometric Dead.

(2)

He the Friend and Father, Francis Galton,
Man of subtle thought and simple speech,
Laboured still when weight of years was on him,
Learning, so that he again might teach.

(3)

Once again the fiery-hearted Weldon
Leaves his birds and beasts and open land,
Seeks the shaded study's dubious twilight,
Worker with our Biometric Band.

(4)

From the prison's gloom emerges Goring,
Seeks the danger-zone in Galton's camp,
Kindest leech and keenest man of science,
Holding high the Biometric Lamp.

(5)

Macdonell...[the name only written here].

Toast Card

NINTH ANNIVERSARY DINNER

BIOMETRIC

&

GALTON LABORATORIES

Tuesday, January 17, 1928

DESSERT IN THE LABORATORY

TOASTS

1. IN PIOUS MEMORY OF SIR FRANCIS GALTON
2. IN REMEMBRANCE OF ALL BENEFACTORS
3. IN MEMORY OF THE BIOMETRIC DEAD
4. THE GUESTS
5. THE POSTGRADUATE AND STUDENT WORKERS
6. THE PROSPERITY OF THE LABORATORY

COFFEE IN THE
COMMON ROOM

HENRY RICHARD HOPE-PINKER

1849–1927

NUR DEM ERNST, DEN KEINE MUHE BLEICHET
RAUSCHT DER WAHRHEIT TIEFVERSTECKTER BORN,
NUR DES MEISSELS SCHWEREN SCHLAG ERWEICHET
SICH DES MARMORS SPRODES KORN. *Schiller*

DIE HOCHSTE WIRKUNG DES GEISTES IST, DEN GEIST HERVORZURUFEN.
Goethe

"Well, here is my last toast to our guests of to-night, and I couple with it the name of Sir Gregory Foster."

The toast-cards with which Pearson provided us at dessert were both a surprise and a delight, with their photographs, sayings or poems; in composing them his fancy would roam freely over past and present. In 1923 a pencil sketch of Francis Galton at 88 made by his niece, Mrs Ellis, no doubt present on that occasion, faced some words of Florence Nightingale's about Mr Galton and Statistics. In 1927 silhouettes of Galton at 8 and at 65 were faced by an extract from John Graunt's *Observations made from the Bills of Mortality* and backed by some lines from George Meredith. The cards of 1926 and 1928 were associated with Weldon and with Hope-Pinker; portions of these are reproduced in Plates VI and VII. The issue of the twenty-first volume of *Biometrika* was commemorated in 1930 with a photograph of the Oxford Darwin statue and some extracts from the Editorial articles written by the three founders of the Journal. In 1931, we had a portrait of H. E. Soper, "1865–1930, Mathematician, Athlete, Inventor, Statistician. Worker in the Biometric Laboratory, 1908–1921"; opposite was an extract from Matthew Arnold's *Thyrsis*:

A fugitive and gracious light he seeks,
Shy to illumine; and I seek it too.
This does not come with houses or with gold,
With place, with honour, and a flattering crew;
'Tis not in the world's market bought and sold—
But the smooth-slipping weeks
Drop by, and leave its seeker still untired,
Out of the heed of mortals he is gone,
He wends unfollowed, he must house alone;
Yet on he fares, by his own heart inspired.

Finally in 1933, at the last dinner, we found waiting in our places a card with a photograph of Miss Elderton who, thirty years before, had become assistant to Galton; this little tribute to one who had done so much to build up the spirit and tradition of the laboratories was something in which we all rejoiced. The card bore, too, in William Watson's words, a farewell message from our host and a note of questioning as to the future.

Guests of the ages, at To-morrow's door
Why shrink we? The long track behind us lies,
The lamps gleam and the music throbs before,
Bidding us enter—and I count him wise,
Who loves so well Man's noble memories,
He needs must love Man's nobler hopes yet more.

Yes, we could carry on into the future, but where was the Toast-master who could succeed K. P.? Such, among many thoughts, were passing through our minds.

We must regret now that one who felt so keenly the need to record in word or portrait the likeness of his friends was so rarely photographed and never satis-

factorily painted. Miss Footner's pencil sketch which stands as the frontispiece of the first part of this memoir was the only good portrait of post-war years; I must, however, refer to another work which resulted from the friendship with Hope-Pinker. Having used Pearson as a model for Roger Bacon before the war, the old sculptor, some ten years later, felt that he would like to try his hand at a portrait bust. He decided to make a direct cut from the marble; it was fascinating to watch the form appearing slowly out of the block, but it was a hard task for a man of over 70 to undertake. After paying many visits to the studio in West Kensington, Pearson at last persuaded Hope-Pinker to bring his marble and tools to University College where he could give him after-lunch sittings in the well-lighted Instrument Room on the second floor of the Laboratory. Even then it was a long and trying occupation, with Pearson at times fretting to be off to his own work, but his affection for the artist, who was so obviously enjoying himself at the job, carried him through some three years of sittings to the end. Once Hope-Pinker brought Tonks over to look at the work, and it was reported that the Head of the Slade School, seeing the large skylight in the room, departed muttering "Ye Gods, top-lighting like this for Science. I must see the Provost!" Poor K. P., poor Hope-Pinker, would authority step in and decree that our rather empty top floor should be handed over to the Slade? But nothing happened and at last the bust was finished and exhibited at the Academy of 1924, without the name of the subject, but bearing the following inscription:

"From life a cut direct but still my friend."

The photograph opposite p 221 does more justice to the craftsman than to his work, but when viewed from a more favourable angle one sees that something of K. P.'s strength has been caught even if the likeness is not altogether a close one; at 75 the chiselling of the stone was a hard job for those old wrists, whatever the spirit behind them, and perhaps at the end they were not content to leave well alone.

It is not I think altogether fanciful to see a link between those hours spent in watching the sculptor shape a head with the help of eye and calliper in his three-dimensioned space, and that novel form of investigation to which Pearson gave much time in the last ten years of his professorship, the study of the relationship between the skulls, or reputed skulls, of certain famous persons and their portraits.

He did not of course regard this research in the same scientific category as his other craniometric work; it was carried out to a large extent as a hobby, but it is worth considering as such for the light it throws on one aspect of his many-sided personality. The first biometric study on these lines, involving the application of laboratory methods to historical inquiry, had been carried out by Miss M. L. Tildesley, her paper on "Sir Thomas Browne: his skull, portraits and ancestry" having been published in *Biometrika* in 1923 (112). In this work both her chief at the Royal College of Surgeons, Sir Arthur Keith, and Pearson who had supplied her cranio-metric training, took much interest. A year later Pearson published a short paper

16/7/24 ll

My dear Dr Bell

In the M.S.S Room at the Br. Mus the other day, they said I must go to the Reading Room to see a certain Indian Ink Drawing of Edinburgh. It occurs in the Catalogue of the Manuscript Maps Charts & Plans Vol II, 1844 under Scotland, Edinburgh. and is entitled:

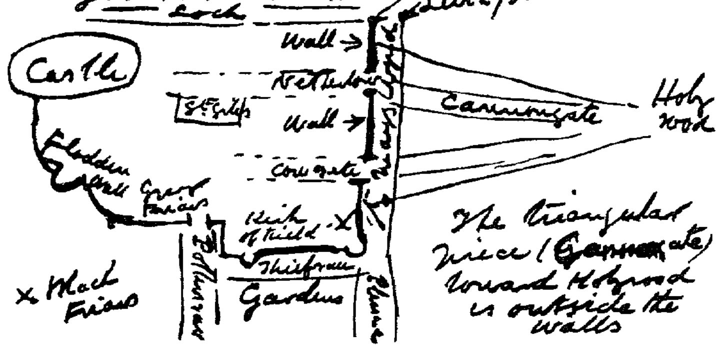
A view of Edinburgh from Hairs. Red Hall in the foreground. Indian Ink by John Clerk. Royal XLIX 2.64. 2.art.64

The question is, does this throw any light on the Kirk of Field & "Flodden Wall." The latter is the southern wall of Edinburgh. Now

Plans or views of Edinburgh are usual taken from the south in which case the Castle is on the left & the Holyrood Palace on the right, or from the north when the castle lies on the right & Holyrood on the left. This is most important as without remembering it east & west & north & south may be interchanged.

South of Edinburgh and therefore in the distance in the first form of plan, and in the foreground of the second lies the Kirk of Field.

In views from the south the general scheme is such



Letter regarding Kirk o' Field of July 16, 1924.

2

The city was bounded by a loch on the north by the east wall with Netherbow & "Cowgate opening into St Mary's Wynd, and by the Flodden Wall from the South East bastion U to the Castle, which wall goes round the Blackfriars & Kirk of Field to the Pottenraw Port & thence up to the Castle. Now the problem is whether John Clerk's plan shows the Kirk of Field, or the "Colledge" afterward built on it. The Kirk of Field may have one or two towers in the centre when with two towers, this

is before 1550, but in the 1567 plan it is as above. If Clerk's plan is from the north you may possibly see the church (to the right of St Mary's Wynd) well up in the distance. What I want is a description of anything in the neighbourhood of the Kirk of Field or in the wall from bastion to Potterraw Port if there be anything.

I enclose a rough sketch I made from a map showing old town & the modern street names.

I am afraid it is rather a thankless job! Most probably Clerk's plan has nothing - than perhaps only an orchard at the Kirk of Field, but I did not like to pass it by without knowing whether it ought to be more thoroughly investigated before I wrote up my Topographical section. Yours very sincerely

Karl Pearson

Letter regarding Kirk o' Field of July 16, 1924 (*continued*)

discussing the cast of a skull reputed to belong to Robert the Bruce (113); besides giving measurements, he fitted the profiles of the skull on to certain fanciful portraits and concluded that the latter were of little value.

"Even the aged dream dreams," he wrote, "and I should like to see a national monument to Bruce at Westminster, an effigy based on the skull as only a great sculptor can conceive it. But it should be the gift of Englishmen only to the united nations It was the Norman element in Bruce, quite as much as the Celtic, which carried him and Scotland to victory at Bannockburn....A great portrait of Bruce is still possible, and if fitly executed would go a long way to justify the value of craniological study for the portraiture of national worthies."

Two years later in 1926 he devoted his Henderson Trust lecture at Edinburgh to an account of the skull and portraits of George Buchanan (114). Before this date he had, however, almost completed a much bigger task dealing with this period of Scottish history, a critical analysis of the events surrounding the murder of Darnley and the light thrown on them by a study of the skull and portraits of that unfortunate youth (115). This work of historical research was carried out by Pearson with extraordinary thoroughness and care of detail; in Plate VIII I have shown a facsimile of a letter which he wrote to Dr Julia Bell* on a point regarding the position of buildings round the Kirk o' Field, the site of Darnley's murder. The immediate problem that he had in mind was whether the markings on Darnley's skull were (*a*) due to syphilis, (*b*) produced by the explosion that was reputed to have caused his death, (*c*) caused by the action of insects or tree-roots after burial. In considering the possibility of (*b*), it was necessary to sift all the available evidence regarding the position of the lodgings where Darnley lay sick and the site where his body was found. In his final conclusions Pearson favoured the hypothesis (*a*), and believed that this fact, if it were true, threw "a flood of light on many points of those dark pages of Scottish history from Darnley's marriage to his murder." The paper was dedicated to the memory of Walter W. Seton, formerly Secretary of University College and Lecturer on Scottish History. How great is the satisfaction when at last we can make an honourable peace with our foes! For many years Seton, in his official capacity as Secretary, and Pearson, an ever-ready champion of freedom against authority, had been again and again at loggerheads. But at last, from their discussions of Mary Stewart over the lunch table, these two widely differing personalities found a bond of interest which broke the barrier. To both, Mary was of all the Stewarts "the most generous, the most cultivated and the most liberal in religion," and both were in agreement that the personal tragedy of her death "was outstripped by a greater tragedy, the strangling of the growth of a national culture and a national spirit by the insatiable greed of rival Tuchuns†." In Moray, Archbishop Hamilton, Bothwell,

* Pearson's habit of retiring to the country to write meant frequent letters of this kind full of queries. Of all his staff he relied most I think for help in this connection on Miss Bell, whose accuracy and skill in following up difficult points had already made her an ideal contributor to *The Treasury of Human Inheritance*.

† I quote from the concluding remarks of Pearson's paper.

Lennox and Morton, even in Elizabeth, Seton and Pearson could now in common find the villains in the piece.

Six years later Pearson returned to this same form of historical research in his joint study with G. M. Morant of the Wilkinson Cromwell head (116).

In following a link between Hope-Pinker, the sculptor, and Pearson's work on portraiture and skulls, I have run on too quickly in time. Before approaching the subject of Darnley, in fact as soon as he had organised the working of his post-war laboratories, Pearson's mind naturally turned back to the unfinished *Life of Francis Galton*. In 1922, the centenary of Galton's birth, he had written a short essay on Galton and his work, published in the *Questions of the Day and of the Fray* Series (117). With this essay we may well associate an appreciation of Charles Darwin, delivered in 1923 as one of a series of London County Council lectures to teachers on "Master Minds of Science," and later published (118). The lecturer gave a charming sketch of Darwin's character and emphasised the effect which *The Origin of Species* had in freeing modern science from the shackles of superstition.

> "I have told you," Pearson said, "that I am young enough to have escaped practically the dogmatic teaching in childhood which placed in bondage the minds of the generations preceding Darwin, and Huxley and Galton. But I am old enough to remember the battles of the sixties and seventies, and the joy we young men then felt when we saw that wretched date B.C. 4004, replaced by a long vista of millions of years of development. Just as much as the older men we looked upon Charles Darwin as our deliverer, the man who had given a new meaning to our life and to the world we inhabited."

The completion of the Galton *Life* was impossible without funds; luckily in 1922 a generous gift from Mr Lewis Haslam, M.P., an old schoolfellow of Pearson's, enabled him to face the difficulties of a second volume. This, the fruit of several vacations' labours, was completed and published in 1924. It dealt with the researches of Galton's middle life in anthropology and heredity, in psychology, in photography and finally his earlier inquiries in the field of statistics.

Pearson was fascinated by the suggestiveness of much of Galton's work belonging to this period; he had studied few of these earlier papers until he met them now, forty years later, in his capacity of biographer. He found many ideas that he would have liked to have been young enough to follow up; applications, for example, of photography in determining types by composite portraiture, in analysing expression and change in the human countenance, in measuring resemblance. But what it was now too late to do himself he hoped that others might some day attempt, if he could set on record here in one book, a résumé of Galton's scattered work, together with an account of his instruments and apparatus, adding as his own contribution the explanatory comment and sympathetic criticism which his personal contact with Galton made possible. There was again to be a halt of some years before the final volume was taken in hand.

During this period, 1923–1929, Pearson was giving a series of lectures on the History of Statistics. It was an extraordinarily interesting course, which started

Outline of Pearson's Lectures on the History of Statistics.

17th century

John Graunt 1620—1674

Edmund Halley 1656—1742
Caspar Neumann 1648—1715

18th century
Abraham de Moivre 1667—1754
Thomas Simpson 1710—1761
W. Kersseboom *c.* 1700—*c.* 1760
George Louis Le Clerc Buffon 1707—1788
Antoine Déparcieux 1705—1768

Colin Maclaurin 1698—1746
James Bernoulli 1654—1705
John Bernoulli 1667—1748
The Newton-Leibnitz Controversy
Leonhard Euler 1707—1783

John Arbuthnot 1667—1735
William Assheton 1641—1711
Assurance Societies founded

Life		Fire:	
Mercers'	1698	Hand in Hand	1696
Amicable	1706	Sun	1706
Union	1714		
Lloyds	1716		
Westminster	1717		
etc.			

John Graunt and Sir William Petty 1623—1687

Gregory King 1656—1714
Charles Davenant 1648—1712

William Derham 1657—1735
Bernard Nieuwentijt 1654—1718
Johann Peter Süssmilch 1707—1767

Nicholas Struyck 1687—1769

Thomas Bayes *c* 1702—1761
Richard Price 1723—1791

French Writers 1750—1800
Marie Jean Antoine Nicolas Caritat de Condorcet 1743—1794
Jean le Rond D'Alembert 1717—1783
Joseph Louis Lagrange 1736—1813
Pierre Simon Laplace 1749—1827

The scheme set out above shows roughly the lines which Pearson followed from John Graunt to Laplace.

with an account of John Graunt and continued, with several interruptions when the lecturer had not time for the necessary preparation, until the period of Laplace had been reached. In the table on page 225 is printed a rough scheme showing the order in which he treated the most important contributors to Statistics in several descending lines. Pearson used to say that he believed that a University teacher ought to give every year one new course on a subject which he had not prepared for lecturing before; only so would he prevent himself from becoming stale. His own contribution to this ideal was certainly given lavishly. With his usual thoroughness he went in every case to the original source for his material; he told us not only of the writings of the individuals selected for discussion, but something of their life history and personality, and to this he added his own comments on the influence which contemporary events had upon them and on the significance of their contributions to the history of the subject. As these lectures were written out by him in very full detail, it is hoped that it will be possible shortly to arrange for their publication.

In 1923 Pearson's eyesight had begun to fail; although this only curtailed certain aspects of his work, it was naturally a cause of much depression. A successful operation for cataract in the summer of 1926, however, removed the fear that his days of active work would be prematurely cut short. But in the same year which eased this tension another blow was to strike him; his wife was laid up with a long illness which only ended with her death in March 1928.

Maria Sharpe sprang, like her husband, from a stock of Dissenters; her family had been notable for its wide interests and independence of thought. William Sharpe, her father, and his brothers had been early brought into touch with their uncle Samuel Rogers, the poet, whose hospitable table had been a centre of literary and artistic fashion during the first half of the nineteenth century. Among her uncles were Samuel Sharpe, the banker and Egyptologist who made and published his own translation of the Bible, Daniel the geologist and admirer of Lyell, and Sutton, the friend of Cuvier, of Stendhal and of Prosper Mérimée. Her father himself had thought of training as an architect before he turned to the law, and before his marriage had been on many holidays abroad, observing and recording with a skilful pencil the towns and countryside of Italy, Switzerland and France. Later in life, with the help of the books in his library, with many illustrations of the old masters and of the leading artists of the day, he was able to convey to his two sons and six daughters much of his own enthusiasm for the reasoned study of literature and art. The scientific renaissance which followed the publication of *The Origin of Species* had its influence, too, on the minds of the young Sharpes as on those of so many of their generation; the message came to them in a variety of ways, in readings of Darwin's *Descent of Man*, in lectures by Huxley on Evolution, by Carey Foster on Electricity, even in "talks to Ladies" on Human Physiology by Elizabeth Garrett Anderson, whose plain speaking shocked some of the elder generation.

To pass from an appreciation of literary and artistic values to a thirst for scientific knowledge involved no big step for minds searching to find in things an

ideal of unity. As Maria Sharpe had written in an article on "Henrik Ibsen; his Men and Women," published in *The Westminster Review* in June 1889:

"We are beginning nowadays to realize the oneness of the laws of life, and we know that in the future the man of science, the man of religion, the moralist, and the social philosopher, equally with the poet and the painter, will take a place in the brotherhood of artists, between whom there is no antagonism, who work by means of observation, selection and imaginative creation to help mankind to make itself."

As I have already mentioned, Maria Sharpe and Karl Pearson had first been brought into touch over the work of the small club whose object had been to encourage scientific investigation of all that concerned "the mutual position and relation of men and women," in past and present. When, soon after their marriage in 1890, Pearson had joined forces first with Weldon and afterwards with Galton, and the science of biometry was born, the wife with such an outlook was able to watch with whole-hearted sympathy her husband's efforts at "observation...and imaginative creation." She was no mathematician and made no attempt to follow the details of his scientific work, but she could appreciate the broad outline of objectives and methods. She saw, too, something of the romantic aspect of those early days of biometry, with the small group of workers fighting for ideals against a critical opposition. There is a long Lay in the metre of Macaulay, but touched by a strong sense of humour, which she wrote after the 1903 holiday meetings of biometricians at Peppard on the Chilterns. I shall quote the two verses in which she referred to the position of Galton's niece, Miss Eva Biggs, of herself and of Mrs Weldon, whose common duty was to guard the peace of the inner circle of biometric workers.

"And round this *inner* circle of learning high and deep,
An *outer* circle ever does watch unceasing keep,
The agile niece, our artist, and one the 'buffer-state'
Who three rampaging urchins doth carefully abate.

"One too, the model hostess, and friend to all of these,
Who has so oft presided at 'Biometric Teas,'
And yet when wheels are driving, and work is to the fore,
Will sort and count and cipher, no inner circler more."

In many ways through her married life, Maria Sharpe Pearson had played nobly the part of the "buffer-state"; with children, with household troubles, with worried laboratory staff, with barking dogs and litters of puppies. To her love and admiration for her husband she added a certain power of detachment; from the days when she was club secretary in the '80's she had seen how his uncompromising search on the track of knowledge and truth had brought him into conflict with other minds and she could give helpful counsel which had eased many situations. The traditions which surrounded her own upbringing had provided her with much of the same philosophy of life as Pearson, to her, also, the moral was the social, the

a-moral the anti-social. Above all she could understand that "creed of life" which makes a man "serve science from love as men in great religious epochs have served the Church," and she knew well how much it was worth while to put aside for its sake.

It was with this companionship that the final link was severed on March 30th, 1928.

With the house and countryside at Coldharbour too full of memories of the past twelve years, Pearson spent the summer vacation of 1928 with his daughters in the Black Forest. Here he was at work on the third and last volume of the *Life of Galton.* The following letter gives some account of their day's programme:

Gasthaus zum Ochsen,
Saig.
August 26, 1928.

My dear Dr Bell,

It was very pleasant last evening to receive a letter from you, and hear of your doings. I fear I cannot make compensation for the letter my wife used to write, for she had a great faculty for writing sympathetic letters. But such as I can, here it is. S—— and I have had a very quiet time, I fear dull for her as we have had little change of thought with any one, but the one or two Germans next us at meals. I have not been very good at expeditions, because I am liable to get overheated and had five days in my room with lumbago as a warning. It is absurd as one grows old to have to take so many precautions. However the weather has been most favourable, and we have been able to sit out a great deal, and do a good deal of writing outdoors. I have completed most of the chapter on Finger-Prints for the Galton Life in this way. Only in my unfortunate manner, I have been led to try and go beyond Galton, which delays matters! Finger-Prints are very fascinating, and the Laboratory ought, I think, to work more at them. Our usual rule is to sit out in the mornings on a more or less distant bench. At least I sit on the bench, and S—— generally lies on the ground. Then dinner at 12.30! After dinner we go for a rather longer walk, reaching some small *Wirtschaft* or inn about 4 o'clock, where we have coffee. At 7.30 we have supper, a rather more frugal meal than *Mittagsessen,* and then we have read Emil Ludwig's *Bismarck* until bedtime at 10 o'clock. It is strange to find now the grandchildren of the old peasants I knew in 1879 and 1880 in possession of their houses. There are very few even of the intermediate generation left. We have met perhaps half-a-dozen men who were prisoners of war in England They speak contentedly of their treatment. Those who were not prisoners are apt to speak bitterly of the occupation of the Rhine. Of the more educated classes, it is not easy to find out their views, except in the newspapers. They are very polite, at times friendly, but you cannot find out what they are really thinking....

I am writing this on a bench, where I can just see the faint outline of the Alps. On a clear day after rain we see the whole chain from Mont Blanc to the Tyrolese Alps, but to-day there is a haze, a soft wind, sun, and the peasants coming back from mass..

I can't recall Kepler's features, but we have two or three, I think, authentic portraits in the Laboratory. Mrs Rollo Russell has given us the 'Nature' series framed, which we must find room for somewhere....

I am rather more surprised at your high fraternal than low avuncular relations. You forget that external blood comes into the stirp, when you deal with collaterals?

I am very glad the Index goes forward, but know full well it is a stiff job!

I expect to be back about the 14th.

Yours very sincerely,

KARL PEARSON.

The year 1929 saw the work on the Galton *Life* finished; its publication in 1930 was made possible by gifts from Miss Dorothy Chase Rowell of Columbia University and Mr Henry Mond. So the big task was at last completed and Pearson could feel that he had left a permanent record of the life and work of the founder of Eugenics; with this he was more concerned than with satisfying any immediate public.

"In the centuries to come," he wrote in the *Preface*, "when the principles of Eugenics shall be commonplaces of social conduct and of politics, men, whatever their race, will desire to know all that is knowable about one of the greatest, perhaps the greatest scientist of the nineteenth century. I have endeavoured to put together many things of which the knowledge in another fifty years will have perished, or not improbably the documents on which that knowledge could be based will be distributed in many directions. I have to the extent of my judgment and powers given an account of Galton's scientific work and of his social ideas, so that all that is essential to an appreciation of his labour and thought will be found in these volumes without the need for continual reference to widely scattered papers, and in the future to still more widely scattered letters."

The main volume, III A, contained three chapters, "Correlation and the Application of Statistics to the Problems of Heredity," "Personal Identification and Description" and "Eugenics as a Creed and the Last Decade of Galton's Life;" in volume III B was a long series of interesting Galton family letters and an Index covering the whole *Life*, prepared by Julia Bell. The long third chapter to volume III A, with its many letters exchanged between the founder and the director of the infant Galton Eugenics Laboratory, provides us with a wealth of information, not only about Galton but about those middle years of Pearson's life; we see him from many angles, in friendship, in controversy, in organisation, in scientific research, in relation to his staff, in holiday mood. The reader of the future will hardly regret the introduction of this autobiographical element.

In August 1930 with the Galton *Life* behind him Pearson was at Saig with Margaret V. Pearson, his second wife, who had for many years been a member of the staff of the Department; a short visit was paid to his old university town of Heidelberg on the way back. There remained three years more before he gave up the helm. They were years in which the number of students who came for training in the Department of Applied Statistics was steadily increasing; years in which he published in *Biometrika* some dozen contributions to statistical theory and in which he spent much time and energy over the completion of two other tasks which he had long had before him, the issue of Part II of *Tables*

for Statisticians and Biometricians (102) and of *Tables of the Incomplete Beta-Function* (103). .

In the summer of 1933 he resigned his professorship; the decision had been taken in July 1932, so that the College and University might have a full year in which to consider the appointment of a successor and to plan any reconstruction of the Department which might appear to them desirable. The following paragraphs which Pearson had included three years before in his last Report to the Court of the Worshipful Company of Drapers would have served too as a final report to the University of London. What of the future, would those with whom lay the power build or destroy?

"My object during the past forty years has been to build up a Laboratory unique of its kind, a place where a novel calculus should be applied to problems concerning living forms. This purpose involved the development of a new form of mathematical analysis, which has grown largely through the work of my pupils scattered through the world, or through those studying their writings. It will continue to grow, but it will only grow with due sense of proportion, if in touch with practical needs, and if it develops in association with anthropometry, medicine, biometry, and the sciences of heredity and psychology. That is to say, if our new calculus is not to become a field for the exploits of the pure mathematician, it must be linked with investigations into topics where its aid is most needed; it must remain a practical science, i.e. applied statistics.

"I have penned this statement in explanation of the manner in which the Laboratory has been built up and expanded. It is, of course, owing to the smallness of its funds, only the framework of a future structure. But that framework I should like to see firmly established before I leave my post. We need readers in anthropometry, biometry and genetics, especially human genetics.

"Such a Laboratory would have seemed a vain dream forty years ago, but we have gone a long way towards it since then. The most remarkable factor in European scientific progress in this direction has been the development in the last ten years of laboratories precisely on these lines—the combination of anthropometry, medicine, and heredity, with a statistical basis—and this development has occurred in a number of European countries. The Laboratories at Lund, at Berlin and Zurich, are built up on the lines of our work here. But they start with many advantages—beyond a knowledge of our experience; they have ample funds, largely provided by Government or university grants, but also by private donors....

"I am writing these pages fully aware that this may form my last Report on the work done here to the Court of the Drapers' Company, and I do so with the full sense of all that Company has done for thirty years to enable me to carry out the aim of my scientific life, the realisation of my dream of forty years ago; but all that aid, and all the work of building up such a laboratory as the present, will have been in vain if the framework is not maintained, and we leave it to other nations to profit by conceptions originating in our own land. I am much more anxious for the permanent establishment of the Laboratory on a sound footing—the completion of its buildings and the permanence of a highly-trained staff—than for my own few remaining years of office. I admit that the progress of the Laboratory in the future will need careful watching and consideration,

but I would hope that the court would pay in the first place attention to what the Laboratory has done and is designed to do, *rather than considering its present director* as by long service entitled in any way to a continuance of the grant If the Laboratory has maintained its reputation for the quality as well as the quantity of its work, it has been in the first place due to the younger generation of trained workers, who have remained faithful to its traditions during the past ten years."

I cannot perhaps do better than close this section with the message Pearson sent through Miss Elderton to his staff and students of that last year on receiving from them an unexpected farewell present of the two large volumes of the Oxford Dictionary:

The Old School House, Coldharbour,
Under Dorking.
September 2nd, 1933.

My dear Dr Elderton,

It was a great surprise to me receiving the Dictionary and your letter this morning. You know through how many years the Laboratory and its Staff have been the greatest of joys to me! Of course it is a very hard task to part with both, but the last two years I had begun to "mind the labour," and felt myself lacking in the requisite energy to cope with obvious difficulties. But the past 22 years will ever be for me the pleasantest of memories, and I hope to keep in touch personally as well as in memory with all old members of the staff. I shall hardly need its delightful present to bear their affections in mind, but they could not have chosen a more serviceable memento or one of greater value to an Editor. Please convey to them all and severally my sense of their goodness and kindness in choice of this keepsake.

Bradshaw once said to me that he held that gifts between real friends should only be flowers—and I added wild flowers. But the world has not reached that standard yet. He once put the fifteen or twenty volumes of Grimm's Dictionary on my table at Cambridge, saying they would be of more use to me than to him; I took it back to his shelves, remarking that dictionaries are not flowers. But I have it now, for I bought it when he died. I will treat your dictionary as the equivalent to a gift of wild flowers in the customary world of to-day.

Always yours sincerely,

KARL PEARSON.

EPILOGUE

The committee appointed by the College decided, after long deliberation, to divide the Department of Applied Statistics into two independent units, a Department of Eugenics with which the Galton Chair would be associated and a Department of Statistics. The existing accommodation, equipment, funds and staff were to be utilised to form these two new Departments. With this scheme Pearson was in serious disagreement. Not only did he feel that the division went against the spirit of all that he had worked for, "a Laboratory, unique of its kind... where a novel calculus should be applied to problems concerning living forms," but he believed that there was almost a breach of trust, since nearly the whole equipment and funds had been supplied in the past for use in such a single

institution. While he agreed wholeheartedly that the College should have a flourishing Department of Statistics which should be primarily devoted to teaching the subject, he could only disapprove of a plan which carved this Department out of the existing research institute, thus limiting the resources available for that establishment. For him, as for Galton, the theory of probability treated from the standpoint of practical statistics was the only sure basis for the study of eugenics, and he feared that this removal of statistics from the Galton Laboratory might result in a future Galton Professor approaching the subject of eugenics from an entirely different basis.

It followed that his retirement from the Department, which must in any case have been a sad event, was made for him more grievous still. He felt that the child of his dreams, this infant laboratory, was to be destroyed by men who had no conception of what it might have grown to. Yet, though the College scheme was approved by the University and a successor appointed to the Galton Chair who was not in sympathy with many of his ideals, though he lived to see his museum broken up and his craniometric laboratory held of small account, with a courage that could triumph over the hardest blows he found joy again in his power still to work. He threw himself into the completion of the monograph on the Cromwell skull and into the editing of *Biometrika*. "When we mind labour, then, then only we're too old."

Was his criticism of the College plan justified and are his fears likely to be fulfilled? That will be for the future to decide. But this can be said now: in the autumn of 1936, about six months after Pearson's death, his old friend Florence Joy Weldon also died, leaving the residue of her estate to found a Chair of Biometry in the University of London, and this has now been established at University College. The founding of this Chair had been a scheme which she had discussed with her husband over thirty years before, in the early days of biometry. Its association with London rather than with Oxford was no doubt due to the existence in the former place of Pearson's own laboratory. Thus, directly or indirectly, from Pearson's election in 1884 to the Professorship of Applied Mathematics it has followed that three new professorships have since been established at University College, those of Eugenics, of Statistics and of Biometry. It will rest largely with the present holders of these Chairs and their successors whether progress is made towards that goal which Pearson had in view, even if the road to be travelled is not precisely that which he had tried to prepare.

October 1933 saw Pearson established in a room placed at his disposal on the other side of the College by D. M. S. Watson, the Professor of Zoology; it saw also R. A. Fisher as the second Galton Professor of National Eugenics and the present writer as head of the new Department of Statistics. The new order had begun.

In his fresh quarters Pearson had round him his books and the most treasured of his pictures, a close array of great figures from the past and friends of more recent years. Elsewhere in the Zoology building was a store-room for *Biometrika*, where Miss F. N. David, his single research assistant, worked. In the many-sided

labour which the issuing of this journal involved, he received from his wife unstinting help. As long as he was able, he kept as formerly his regular College hours and the College terms, this was the discipline of his science.

The most interesting piece of work which he completed in his retirement was undoubtedly the joint paper with G. M. Morant on the Wilkinson Cromwell Head (**116**).

"So much has been written about this Head," the authors stated in their introductory section, "and the controversy has been so keen, that it might appear that there was nothing to be said on the topic which had not been said already. In other words that the authenticity of the Head must be ever left in that state of doubt in which historians and critics have enveloped it. Yet when one has studied the innumerable notes, letters, and newspaper articles one finds only a mass of contradictory *opinions*, repetition of various absurd myths about Cromwell's body, not one single trustworthy measurement or fitting of the head to any form of portrait; in fact the whole of the century of discussion is *vox et praeterea nihil.* Had the authors of the present paper merely wished to contribute surmises, criticisms of earlier surmises, vague statements that the Head was in their opinion like or unlike Cromwell's portraits, there would have been no excuse for this monograph. The essential difference between this and earlier discussions of the subject is (*a*) that the authors have no bias for or against the authenticity; (*b*) that they trust solely to measurements on the Head, and to its good or bad fit to portraits; and (*c*) what has been essential to their investigation, that two great privileges have been granted to them by the owner, Canon Horace Wilkinson, (i) to retain the Head adequately long in order to carry on the comparison with busts, masks and portraits, and (ii) to state freely what conclusions they have reached as a result of their investigations."

The investigation was of a kind which appealed to Pearson immensely. "Don't chatter, make trial," Charles II is reported to have said, and though the "trial" which the Merry Monarch may have contemplated was no doubt less exact than Pearson would have approved, the motto served; the more persons who had merely talked about the Cromwell head, the more eager he was to get down to exact measurement himself! The paper followed the lines of the Darnley inquiry, but it covered even more ground. Its authors intended to enjoy themselves thoroughly and anyone who turned over with Pearson the pages of the album in which the first proofs of the 100 illustrative plates were pasted can have had little doubt of the pleasure which the chief author was drawing from his work. The skill and enthusiasm of the historical investigator who, long before, had collected the Veronica portraits of Christ and the woodcuts of Albrecht Dürer were combined with the perfected technique of that old and experienced measurer of heads. And what was the conclusion? Not expressed in terms of a precise probability measure, but approaching that as nearly as possible:

"The defective history of the Head hinders the *demonstration* that it is Cromwell's, but many a man has been hanged on a smaller amount of circumstantial evidence for his crime than exists for the identity in this case. The probability for the identity is so convincing that any critic need not be considered who cannot produce a higher

probability that this Head must be that of another embalmed and decapitated person of the seventeenth century. Who was he, and do his busts or portraits fit to a higher degree this Head?

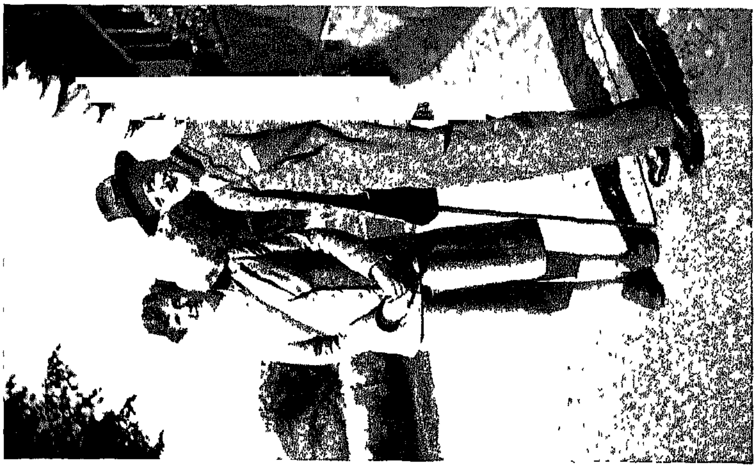
"We started this inquiry in an agnostic frame of mind, tinged only by scepticism as to whether the positive statements made in the past with regard to it were not based solely on impressions unjustified by any attempt at a scientific investigation. We finish our inquiry with the conclusion that it is a 'moral certainty' drawn from the circumstantial evidence that the Wilkinson Head is the genuine head of Oliver Cromwell, Protector of the Commonwealth."

In the mathematical field also Pearson was still active. Continuing his habit of adding to the supply of readily available tables, he reissued photographically Legendre's *Tables of the Complete and Incomplete Elliptic Integrals* (119), prefacing them by an introduction of his own, explaining the tables and their use. He also put in train the computation of a table he had long planned of the probability integral of the correlation coefficient*. What a part the conception symbolised by that little letter r (which might so easily have been c) had played for forty years in the pages of his statistical contributions!

His paper of 1933 (99) on the new goodness of fit test was followed by another of 1934 containing further applications (120). Finally, we may note two letters to *Nature* (121) and a last contribution to *Biometrika* (122) on a problem which, if the word is understood in its widest sense, may be termed the problem of graduation. Here he sought again to emphasise the difference between the world of concepts and the world of perceptual experience. It is the teaching of *The Grammar of Science*, most clearly seen in the letters, but to be read, too, behind the thrusts of the *Biometrika* article. The mathematical equations of frequency curves and regression lines, the probability distributions of estimates and test criteria, the principles of estimation and of testing hypotheses, all these are abstract concepts the application of which to experience involves a process of graduation. He sensed a danger that statisticians might be carried away by the fascination of ideas into attributing some magic significance to these conceptual models, into giving a false reality to words and phrases whose seeming importance was perhaps enhanced by the addition of capital letters, Efficiency, Power, Information, Likelihood. To him the value of these notions could only lie in their utility in the perceptual world; here there could be no ultimate right or wrong, only more useful and less useful, and even then, what was of greater aid to one man might well be of less to another. Such, I think, was the final statistical message that Pearson left us.

In vacation time he was largely at Coldharbour, where he would still walk his 10 or 12 miles over the hills through Friday Street, Abinger or Holmbury St Mary to the farther end of the ridge. In the last three summers he also returned with his wife to the Yorkshire dales, from which his ancestors "had ridden south over the moor," and spent some weeks at Danby, a place of so many memories. The

* This work has just been completed by Miss F. N. David.



At Saig, with his younger daughter, 1928.

In the study at Hampstead, 1933.

Blakey Ridge, Danby, 1935

long July evenings of the north, the good smell of turves and bracken and young heather made the worries of London of so much less significance. As he had written to Weldon long before on July 1st, 1900, at the end of a hard College year:

"I breathe pure air again and feel human once more! My forebears were yeomen on these uplands and my great great grandfather was a fool ever to leave them! I should have known all about inheritance and never wanted to give expression to it in a law had he only stuck to the soil! Now I don't possess a square foot of Mother Earth, and can't carry out a single breeding experiment! Just fancy what a flock of 300 sheep would mean!! I have to beg land and labour of friends for a poppy-patch! My Father sold his ancestral patch last year, because we agreed a man ought to farm his own land and it had been let for years to tenants. I had a sort of vague dream of turning it into a breeding farm, but I never quite realised how it was to be kept going. My Father, being one generation nearer the plough than I, considers himself much my superior, but I never heard anything come of his theories or practices in agriculture, while I am certain nothing would come of mine! Still I long to retire to a plot of land and breed something. Pigs or sheep,—poppies or snails are all one to me. Even house-sparrows would be exciting, for I should dearly like to know if egg mottling is hereditary!

"I have written some 35 letters since I arrived here on Friday night,—the arrears of correspondence since Easter, and this is the first letter I don't feel it a nuisance to write and therefore you must pardon my chattering. Since writing so far I have dined and strolled out on to the moor at the back. It is 9 o'clock and broad daylight, and the plover quite active and the grouse whirring off, and nothing when you get over the brow of the hill, but miles of blackish heather, scarcely yet in bud, with the green of the young bilberry shoots, and one gray stone ruin—the bell-house—on the causeway which runs for miles across the moor, which some ancestor of mine built and where he doubtless rung the bell for a guide to the mule-drivers taking wool north and south, when the North Sea 'rauk' came upon them. Then turning to the brow again and looking right up Danby Dale with moor on either ridge and a narrow cultivated strip up the bottom, I see ten miles up the Dale to the Quaker's Way, where in the memory of man, they used to come down riding pillion to the meeting on Sunday, long rides 10 or 15 miles over the moor. Those three little garths like potato patches, each about three miles apart up the dale are really stoneless burial grounds and my forebears lie in them and the natives know neither their name or purpose. The ruin by the middle one more than 200 years ago was Hartus' farm, and George Fox preached there, and my 5th great grandparents were married there, Quaker fashion. Opposite it is Lumley House whence. "

and here the page is lost. Perhaps his eye passed on westwards across the dale to Stormy Hall, with its clump of trees, and Honey Bee Nest, up the track to St Helena, the highest farm of all, and so along the ridge to Blakey, to the hut which had once been a Meeting House on the moor whence in the 1680's the soldiers had taken away to York gaol his ancestors Gregory Pearson and George Unthank, steadfast in their Quaker faith. Or from Lumley House he may have passed to the church in the centre of the dale and up over Danby High-moor to the Fryups, by that track, marked by its tall stones, over which within living memory they had still borne by hand the coffins for burial; and so to the ruins of

Danby Castle, once a manor of the ancestors of the Bruce, and back to the village by the stone causeway where the pack mules had passed and where Tommy Pearson, like Tam o' Shanter, had once joined issue with a witch.

In 1935 it became clear that Pearson's strength was gradually failing; that old hard-worked body was at last worn out though the mind and spirit were still eager to carry on. To the last he worked at *Biometrika*, and he had almost seen the final proofs of the first half of Volume XXVIII through the press when he died. The end came at Coldharbour on April 27th, 1936, when spring in those Surrey hills was at its best.

At the funeral service in London a few days later we heard the words, never perhaps more appropriately spoken, which he had more than once applied to others from his favourite Browning's "A Grammarian's Funeral":

> This man decided not to Live but Know—
>   Bury this man there?
> Here—here's his place, where meteors shoot, clouds form,
>   Lightnings are loosened,
> Stars come and go! Let joy break with the storm,
>   Peace let the dew send!
> Lofty designs must close in like effects:
>   Loftily lying,
> Leave him—still loftier than the world suspects,
>   Living and dying.

And the music of the second movement of Beethoven's 7th Symphony which followed, told us in its magnificence something of joy breaking with the storm.

In the course of this survey we have seen Pearson from many aspects, as the historian, the writer on folklore, the socialist, the applied mathematician who discussed problems of elasticity and engineering and theories of atomic structure, as the author of *The Grammar of Science*, as the biometrician, statistician and eugenist, as the teacher and the biographer. It would be hard to say which of his contributions to science are of most importance; the influence that he has had does not depend on any sharply defined discoveries. It would be easy, too, to say that here he made mistakes, there his contemporaries or successors in the light of new facts have, in his own terms, found a simpler logical construct than he had, in which to gather the known phenomena of perception. But in such criticism small profit lies. It is well to recall the words he wrote himself on the influence that great minds have had on the generations which followed them:

"The little men say there was evolution before Darwin; the little men say somebody discovered logarithms before Napier, the belittlers believe that the law of the inverse square was propounded before Newton, and that somebody conceived of Eugenics before Galton. Well, the imagination of man has always run riot, but to imagine a thing is not meritorious, unless we demonstrate its reasonableness by the laborious process of studying how it fits experience, or make it a real factor of practice. Darwin did bring the ideas of evolution home to science; logarithms did come into general use after the

publication of Napier's *Logarithmorum canonis descriptio* (1614); Newton did predict the motion of the moon on the basis of his law of gravitation, and the name and idea of a science of Eugenics have become worldwide only since Galton made his appeal and showed its possibilities....

"The little men say that relativity has killed Newtonian mechanics, but they do not add that now and for long years to come satisfactory answers to ninety-nine per cent. of mechanical and physical problems—problems now essential to our daily existence—will be reached by Newtonian approximations...

"Do these statements belittle Einstein? On the contrary the writer believes that while relativity now modifies the treatment of a very small percentage of physical problems, it will in future modify the treatment of more and more. It is a question of the growth in accuracy of our instruments and the developing refinement of our observational powers. The fundamental importance of relativity at the present time is the manner in which it is changing and must change our attitude towards the physical universe....New phases of philosophy, new phases of religion will grow up to replace the old. But the cultivated mind can never regard life and its environment in the same way as men did before those days of Darwin and before these days of Einstein. The 'value' of words, the 'atmosphere' of our conceptual notions of phenomena, has been for ever changed by the movement which began with Darwin and at present culminates in Einstein." ((117) pp. 5—7.)

Can we not perhaps say that in similar manner, by the long process of studying how it fits experience, Pearson has made the calculus of mathematical statistics a real factor of practice in vast fields of scientific inquiry? Not only did he display the motto, "Until the phenomena of any branch of knowledge have been submitted to measurement and number it cannot assume the status and dignity of a science," but having himself provided a mathematical technique and a system of auxiliary tables, by ceaseless illustration in all manner of problems he at last convinced his contemporaries that the employment of this novel calculus was a practical proposition. From this has resulted a permanent change which will last, whatever formulae, whatever details of method, whatever new conceptions of probability may be employed by coming generations in the future. And if in Pearson's work a critical eye can find here and there a blunder in algebra or arithmetic, or an apparent lack of clearness in thought, it will do no harm to remember his own statement in *The Ethic of Freethought*:

"Every freethinker, then, owes an intense debt of gratitude to the past; he is necessarily full of reverence for the men who have preceded him; their struggles, their failures and their successes, taken as a whole, have given him the great mass of his knowledge. Hence it is that he feels sympathy even with the very failures, the false steps of the men of the past. He never forgets what he owes to every stage of past mental development."

In the spirit of these words let us leave him, paying reverence to a great man who has preceded us, and confident that wherever the path of science may lead, Karl Pearson has contributed his full share of that pioneer work from which alone true progress can follow.

BIBLIOGRAPHY. (PART II)

The following list relates to Karl Pearson's published work which has been specially referred to in this second half of the memoir. The numbers follow on from those given in the first half of the memoir, and as before the order is that in which the items have been mentioned in the text. In the case of papers written by Pearson alone, e.g. (**45**), no statement of authorship is made, while for joint papers, e.g. (**46**), the authors' names are given. For convenience, a few other memoirs are included in the list, e.g. (**55**), which were not written by Pearson but form part of a scheme of investigation which he was either contributing to or planning; these are indicated by the symbol †. It is hoped that a complete Bibliography of Pearson's work prepared by Dr G. M. Morant will be issued shortly as a *Biometrika* publication.

45. "On the Mathematical Theory of Errors of Judgment, with Special Reference to the Personal Equation." *Phil. Trans. Roy Soc.* CXCVIII, A (1902), pp. 235—299.

46. "Mathematical Contributions to the Theory of Evolution. XV A Mathematical Theory of Random Migration," by Karl Pearson with the assistance of John Blakeman. *Drapers' Company Research Memoirs*, Biometric Series, III (1906).

47. "Mathematical Contributions to the Theory of Evolution. XVI. On Further Methods of Determining Correlation." *Drapers' Company Research Memoirs*, Biometric Series, IV (1907).

48. "Mathematical Contributions to the Theory of Evolution. XVIII. On a Novel Method of Regarding the Association of two Variates classed solely in Alternative Categories." *Drapers' Company Research Memoirs*, Biometric Series, VII (1912).

49. "Albinism in Man," by Karl Pearson, E. Nettleship and C. H. Usher. *Drapers' Company Research Memoirs*, Biometric Series, VI (1911), VIII (1913), IX (1913)

50. "A Study of the Long Bones of the English Skeleton," by Karl Pearson and Julia Bell. *Drapers' Company Research Memoirs*, Biometric Series, X (1917–1919), XI (1919).

51. "On the Sesamoids of the Knee-Joint," by Karl Pearson and Adelaide G. Davin. *Biometrika*, XIII (1921), pp. 133—175, 350—400 and an Addendum in *Biometrika*, XIV (1923), p 409

52. Papers of No. 51. Reprinted as *Drapers' Company Research Memoirs*, Biometric Series, XII (1922).

53. "A First Study of the Statistics of Pulmonary Tuberculosis." *Drapers' Company Research Memoirs*, Studies in National Deterioration, II (1907).

54. "A Second Study of the Statistics of Pulmonary Tuberculosis. Marital Infection," by Ernest G. Pope and Karl Pearson. With an Appendix on Assortative Mating, by Ethel M. Elderton. *Drapers' Company Research Memoirs*, Studies in National Deterioration, III (1908).

55.† "A Third Study of the Statistics of Pulmonary Tuberculosis. The Mortality of the Tuberculous and Sanatorium Treatment," by W. Palin Elderton and Sidney J. Perry. *Drapers' Company Research Memoirs*, Studies in National Deterioration, VI (1910).

56.† "A Fourth Study of the Statistics of Pulmonary Tuberculosis. The Mortality of the Tuberculous: Sanatorium and Tuberculin Treatment," by W. Palin Elderton and Sidney J. Perry. *Drapers' Company Research Memoirs*, Studies in National Deterioration, VIII (1913).

57. "A First Study of the Influence of Parental Alcoholism on the Physique and Ability of the Offspring," by Ethel M. Elderton, assisted by Karl Pearson. *Eugenics Laboratory Memoirs*, X (1910).

58. "A Second Study of the Influence of Parental Alcoholism on the Physique and Ability of the Offspring," by Karl Pearson and Ethel M. Elderton. *Eugenics Laboratory Memoirs*, XIII (1910).

59. "A Preliminary Study of Extreme Alcoholism in Adults," by Amy Barrington and Karl Pearson, assisted by David Heron. *Eugenics Laboratory Memoirs*, XIV (1910)

60.† "A Second Study of Extreme Alcoholism in Adults With special reference to the Home-Office Inebriate Reformatory data," by David Heron. *Eugenics Laboratory Memoirs*, XVII (1912).

61. "Treasury of Human Inheritance." Vols I, II and III Part I. Edited by Karl Pearson. *Eugenics Laboratory Memoirs*, VI (1909), IX (1909), XI (1910), XII (1911), XV (1912), XVI (1912), XXI (1922), XXII (1925), XXIII (1926), XXIV (1928), XXVI (1931), XXVII (1932), XXVIII (1933).

62. "Albinism in Dogs," by Karl Pearson and C H. Usher *Biometrika*, XXI (1929), pp. 144—163.

63.† "On the Inheritance of the Diathesis of Phthisis and Insanity. A Statistical Study based upon the Family History of 1500 Criminals," by Charles Goring. *Drapers' Company Research Memoirs*, Studies in National Deterioration, V (1909).

64. "Tuberculosis, Heredity and Environment" *Eugenics Laboratory Lecture Series*, VIII (1912).

65. "The Fight against Tuberculosis and the Death-rate from Phthisis." *Questions of the Day and of the Fray*, IV (1911).

66. "Supplement to the Memoir entitled 'The Influence of Parental Alcoholism on the Physique and Ability of the Offspring.' A Reply to the Cambridge Economists." *Questions of the Day and of the Fray*, I (1910)

67. "An Attempt to correct some of the Mis-statements made by Sir Victor Horsley, F.R.S., F.R C.S., and Mary D. Sturge, M.D , in their Criticisms of the Memoir. 'A First Study of the Influence of Parental Alcoholism, &c.'" *Questions of the Day and of the Fray*, III (1911).

68.† "A First Study of the Statistics of Insanity and the Inheritance of the Insane Diathesis," by David Heron *Eugenics Laboratory Memoirs*, II (1907).

69.† "Mendelism and the Problem of Mental Defect. I. A Criticism of Recent American Work," by David Heron. *Questions of the Day and of the Fray*, VII (1913).

70. "Mendelism and the Problem of Mental Defect. II. On the Continuity of Mental Defect," by Karl Pearson and Gustav A. Jaederholm *Questions of the Day and of the Fray*, VIII (1914)

71. "Mendelism and the Problem of Mental Defect III. On the Graduated Character of Mental Defect and on the need for standardizing Judgments as to the Grade of Social Inefficiency which shall involve Segregation." *Questions of the Day and of the Fray*, IX (1914).

72. "A Second Study of the Variation and Correlation of the Human Skull, with Special Reference to the Naqada Crania," by Cicely D. Fawcett, assisted by Alice Lee and others. *Biometrika*, I (1902), pp 408—467

73. "Further Considerations on the Correlation of Stellar Characters," by Winifred Gibson and Karl Pearson. *Monthly Notices of the Royal Astronomical Society*, LXVIII (1908), pp 415—448

74. "On Some Points with regard to the Light-Fluctuation of Variable Stars," by Karl Pearson, assisted by Julia Bell. *Monthly Notices of the Royal Astronomical Society*, LXIX (1908), pp. 128—151.

75. "On a New Method of Determining Correlation between a Measured Character A, and a Character B, of which only the Percentage of Cases wherein B exceeds (or falls short of) a given Intensity is Recorded for each Grade of A." *Biometrika*, VII (1909), pp. 96—105

76. "On a New Method of Determining Correlation when one Variable is given by Alternative and the other by Multiple Categories." *Biometrika*, VII (1910), pp. 248—257.

77. "On the Measurement of the Influence of 'Broad Categories' on Correlation." *Biometrika*, IX (1913), pp 116—139.

78. "On the Influence of Past Experience on Future Expectation." *Philosophical Magazine*, XIII, 6th series (1907), pp. 365—378.

79. "On the Probability that two Independent Distributions of Frequency are really Samples from the same Population." *Biometrika*, VIII (1911), pp. 250—254.

80. "On the Correlation of Fertility with Social Value," a Co-operative Study. *Eugenics Laboratory Memoirs*, XVIII (1913).

81.† "Report on the English Birthrate. Part I. England, North of the Humber," by Ethel M. Elderton. *Eugenics Laboratory Memoirs*, XIX, XX (1914).

82. "The Problem of Alien Immigration into Great Britain, illustrated by an Examination of Russian and Polish Jewish Children," by Karl Pearson and Margaret Moul. *Annals of Eugenics*, I (1925), pp. 5—127; II (1927), pp. 111—244, 290—317; III (1928), pp. 1—76, 201—264

83.† "On the Relative Value of the Factors which influence Infant Welfare," an Inquiry by Ethel M. Elderton based on data provided by Medical Officers of Health. *Annals of Eugenics*, I (1925), pp. 139—243, 277—381; III (1928), pp. 96—193. These papers were afterwards reprinted together as *Eugenics Laboratory Memoirs*, XXV (1928).

84. "On the Handicapping of the First-Born." *Eugenics Laboratory Lecture Series*, X (1914).

85. "The Relative Strength of Nurture and Nature," by Ethel M. Elderton. *Eugenics Laboratory Lecture Series*, III, 1st edition (1909), 2nd edition (1915), enlarged; Part I. "The Relative Strength of Nurture and Nature," by Ethel M. Elderton. Part II "Some Recent Misinterpretations of the Problem of Nurture and Nature," by Karl Pearson.

86. "On the Probable Error of Frequency Constants." Part II. *Biometrika*, IX (1913), pp. 1—10.

87. "On the Probable Error of a Coefficient of Correlation as Found from a Fourfold Table." *Biometrika*, IX (1913), pp. 22—27.

88. *Tables for Statisticians and Biometricians.* Part I. Edited by Karl Pearson. Cambridge University Press (1914); Biometric Laboratory, University College (1924), (1930).

89. "On the Distributions of the Standard Deviations of Small Samples." *Biometrika*, X (1915), pp. 522—529.

90. "On the Distribution of the Correlation Coefficient in Small Samples." A Co-operative Study by H. E. Soper, A. W. Young, B. M. Cave, A. Lee and K. Pearson. *Biometrika*, XI (1917), pp. 328—413.

91. "On the Probable Error of a Coefficient of Mean Square Contingency." *Biometrika*, X (1915), pp. 570—573.

92. "On the Probable Error of a Coefficient of Contingency without Approximation," by Andrew W. Young and Karl Pearson. *Biometrika*, XI (1916), pp. 215—230.

93. "Peccavimus!" Editorial, *Biometrika*, XII (1919), pp. 259—281.

94. "On the General Theory of Multiple Contingency with special reference to Partial Contingency." *Biometrika*, XI (1916), pp. 145—158.

95. "On the Application of 'Goodness of Fit' Tables to test Regression Curves and Theoretical Curves used to describe Observational or Experimental Data." *Biometrika*, XI (1916), pp. 239—261.

96. "Side Lights on the Evolution of Man." *Eugenics Laboratory Lecture Series*, XIII (1921).

97. "The Fundamental Problem of Practical Statistics." *Biometrika*, XIII (1920), pp. 1—16.

98. "Note on Bayes' Theorem." *Biometrika*, XVI (1924), pp. 190—193.

99. "On a Method of Determining whether a Sample of Size n supposed to have been drawn from a Parent Population having a known Probability Integral has probably been drawn at Random." *Biometrika*, XXV (1933), pp. 379—410.

100.† "Tables of the Digamma and Trigamma Functions," by Eleanor Pairman. *Tracts for Computers*, I (1919).

101. *Tables of the Incomplete Gamma-Function.* Edited with an Introduction by Karl Pearson. H.M. Stationery Office (1922).

102. *Tables for Statisticians and Biometricians.* Part II. Edited by Karl Pearson. Biometric Laboratory, University College (1931).

103. *Tables of the Incomplete Beta-Function.* Edited by Karl Pearson. *Biometrika* Office (1934).
104. "On the Biometric Constants of the Human Skull," by Karl Pearson and Adelaide G. Davin. *Biometrika*, XVI (1924), pp. 328—363.
105. "On the Coefficient of Racial Likeness." *Biometrika*, XVIII (1926), pp. 105—117.
106. "The Application of the Coefficient of Racial Likeness to test the Character of Samples." *Biometrika*, XX B (1928), pp. 294—300. "Note on Standardisation of Method of using the Coefficient of Racial Likeness." *Ibid.* pp 376—378.
107. "The Cranial Coordinatograph, the Standard Planes of the Skull, and the value of Cartesian Geometry to the Craniologist, with some Illustrations of the uses of the New Method." *Biometrika*, XXV (1933), pp. 217—253.
108. "Further Investigation of the Morphometric Characters of the Individual Bones of the Human Skull," by Karl Pearson and T. L. Woo. *Biometrika*, XXVII (1935), pp. 424—465.
109. "On our Present Knowledge of the Relationship of Mind and Body." *Annals of Eugenics*, I (1926), pp. 382—406.
110. "On a New Theory of Progressive Evolution." *Annals of Eugenics*, IV (1930), pp. 1—40.
111. "On the Inheritance of Mental Disease." *Annals of Eugenics*, IV (1931), pp. 362—380.
112.† "The Skull and Portraits of Sir Thomas Browne," by Miriam Tildesley. *Biometrika*, XV (1923), pp. 1—76.
113. "The Skull and Portraits of King Robert the Bruce." *Biometrika*, XVI (1924), pp. 253—272.
114. "On the Skull and Portraits of George Buchanan." *Biometrika*, XVIII (1926), pp. 233—256. (The fifth Henderson Trust Lecture.)
115. "The Skull and Portraits of Henry Stewart, Lord Darnley, and their Bearing on the Tragedy of Mary, Queen of Scots." *Biometrika*, XX B (1928), pp. 1—104.
116. "The Wilkinson Head of Oliver Cromwell and its Relationship to Busts, Masks and Painted Portraits," by Karl Pearson and G. M. Morant. *Biometrika*, XXVI (1934), pp. 269—378.
117. "Francis Galton. A Centenary Appreciation." *Questions of the Day and of the Fray*, XI (1922).
118. "Charles Darwin, 1809-1882. An Appreciation." *Questions of the Day and of the Fray*, XII (1923).
119. *Tables of the Complete and Incomplete Elliptic Integrals.* Reissued from Legendre's *Traité des Fonctions Elliptiques*, Paris, 1825, with an Introduction by Karl Pearson. *Biometrika* Office (1934).
120. "On a New Method of Determining 'Goodness of Fit.'" *Biometrika*, XXVI (1934), pp. 425—442.
121. "Statistical Tests." Two letters. *Nature*, CXXXVI (1935), pp. 296 and 550.
122. "Method of Moments and Method of Maximum Likelihood." *Biometrika*, XXVIII (1936), pp. 34—59.

APPENDIX III

EXTRACT FROM KARL PEARSON'S REPORT TO THE WORSHIPFUL COMPANY OF DRAPERS MADE IN FEBRUARY 1918

War Work of the Biometric Laboratory

Since the Report of 1913 made to the Court of the Worshipful Company of Drapers on the work undertaken by aid of their grant, the war has given a wholly different course to the life of the Laboratory and its staff. In July, 1914, we fully expected the main work of the next six months to be the occupation and equipment of the new Laboratory buildings, the fitting up of the public museum and the anthropometric laboratory. All this development and the extension of the biometric work which would have followed it were shattered by the war. The new Laboratory buildings were taken

over by the Government as a military hospital and will presumably be used as such till the end of the war. Very early in the war several members of the staff went off on special war duties for which their training in computing largely fitted them. Of those who in July, 1914, were at work in the Laboratory, Mr Soper left to do experimental work on electrical apparatus for war purposes; Mr Everitt left to train women in the polishing of prisms and lenses for periscopes, etc.; Miss B. M. Cave went as a computer to the Admiralty for naval air-plane work; Dr Heron left as statistical adviser to a large insurance company, of which he has since become secretary; he has further acted as a statistical adviser to the Ministry of National Service. Amongst those who filled the gaps thus arising, Mr Horwitz has since gone as a statistician to the Ministry of Munitions, Mr Firth in a like capacity to the Contracts Department, War Office, and he has recently been followed by Mr Frobisher, the Crewdson-Benington Student, who felt it his duty to ask for the suspension of his studentship that he might undertake similar work under the War Office. In all these cases age or physique disqualified for active military service.

These matters are cited as illustrations of the difficulty at the present time of holding together for pure research work a highly trained staff. Posts could have been found in Government offices at the present time for double the number of my total staff, and many old students and past assistants of the Department are at present employed in one form or another of statistical work, often of a very important or confidential character, for the War Office or the various new Ministries.

The position therefore was at the beginning of the war an extremely difficult one. It was essential for the future to retain if possible a highly trained staff, but the funds at our disposal neither enabled us to compete with the high salaries offered to competent statisticians, nor, if they had been, would it have appeared justifiable to keep members of the staff, who were urgently needed for national work of importance. The only reasonable solution of the difficulty seemed to be the voluntary employment of the Laboratory as a whole on war work, and this in some form wherein its training and computing experience would be of essential value at the present crisis. The feeling that the staff as a whole were doing national work would maintain its *esprit de corps* and retain its more loyal members at their posts, even if more highly paid appointments were proposed to them. Accordingly I discussed the matter with the staff in the first week of August, 1914, and its members agreed to dispense with the best part of their holidays, and to devote their time to war work. With hardly an exception this attitude has been maintained throughout the whole period of the war up to date by my old staff. They have worked to the full extent of their powers and sometimes beyond them, holidays have been few and far between and only taken when some rest was a necessity. I cannot speak too highly of the loyalty and energy of my assistants. In 1916 the Laboratory was kept going throughout the whole of Easter, and for the men the hours have been 9 to 6, Sundays only excepted.

In August, 1914, we started with work for the Board of Trade, Labour Department, the question of unemployment being then a vital one. We prepared fortnightly labour charts, showing the state of unemployment both for insured labour and uninsured labour in all English, Irish and Scottish towns of over 20,000 inhabitants, and in all county districts. Some 600 charts were prepared for each issue and brought up to date. These

were used by the Central Relief Committee for the control of local conditions. By July, 1915, our charts themselves showed that the possibility of great labour difficulties which had been so marked in December, 1914, had vanished, and accordingly these charts ceased. We next worked for the Board of Trade, Census of Production Department, preparing charts of the tonnage used for each type of import at each season of the year with a view to aiding the special officials who had the task of controlling the amount and character of imports and the shipping to be used for them. These charts were followed by a series of charts of the rates of exchange in all the great European and North and South American cities. These were photographically reproduced and kept up to date for a whole year, being distributed by the Board of Trade to various Government offices. Meanwhile more urgent problems had come to us from other sources. During a good deal of 1916 we were occupied with theory and computations concerning torsional strain in the blades of air-plane propellers for a department of the Royal Aircraft Factory at Farnborough. A report was also prepared on the elastic constants of wood for the head of a section of the Admiralty Air Department. In July, 1916, our energies were directed by a member of the same department to bomb trajectories, and several months were devoted to the calculation of such trajectories for use in the sights of bombing air-planes. Our tables have been privately printed by the Air Committee. This work was later extended to combined air and water bomb trajectories in which a number of new problems arise, and we endeavoured to consider them experimentally by using models. The successful solution of these problems would be of great importance in the anti-submarine campaign.

Thus our work during 1916 gradually turned to the more military side of war work. On January 1st, 1917, we were asked to assist Captain A. V. Hill, of the Anti-Aircraft Experimental Section of H.M.S. Excellent, with gunnery computations for anti-aircraft guns, and from that date onwards we have been engaged without cessation in computing ballistic charts and range tables for the Ordnance Committee. We have had in charge the preparation of the whole of the charts and high-angled range tables for the anti-aircraft guns in both Army and Navy, and the preparation of the data for the sights of these guns. All the organisation and control of the work, all the finer draughtsmanship of the charts, was undertaken by the trained members of my staff. The charts have been engraved by the Ordnance Survey and now number twenty. The high-angled range tables are printed at Woolwich, and both charts and range tables are now issued officially by the Ordnance Committee for about a dozen anti-aircraft guns. The work has been so urgent and of such value that the Ministry of Munitions has placed eight to ten computers and draughtsmen at my disposal, and with the exception of one week at Christmas the Laboratory was never closed from January 1st, 1917, to January 1st, 1918.

The main feature of the work, however, has been the voluntary work of direction and control exercised by my staff. Only recently, owing to the rise in prices and the small stipends paid to academic workers, have *honoraria* for holidays and overtime been paid to the junior members of the staff. Voluntary enthusiasm has been the mainspring of the whole enterprise. I venture to think that we may claim it as good evidence of the value of the training given in the Laboratory that our members could thus take upon themselves an entirely new field of work. It must be remembered that high-angled range tables for anti-aircraft guns were unknown before the war had developed the

aeroplane as a new instrument of warfare, that in the great bulk of cases we had to develop new methods from the very rough processes originally suggested to us, and that the authorities at Woolwich have themselves consulted us as to methods of calculation and instruments used. In the course of the work no fewer than 85 new tables have been computed for various guns, giving ranges, fuzes or sights, and, further, several of the old ballistic tables have been re-calculated or developed to higher degrees of accuracy. We have heard from men at the Front due appreciation of our charts and tables, and whereas before their existence practically no air-planes were shot down, we now hear of upwards of sixty in six months by direct anti-aircraft gunfire.

Just before Christmas an urgent demand came from the Front to General Bingham for a remedy against the low-flying German air-planes, which were making things "unhealthy" for our men in the trenches. It was a source of great satisfaction to us, and a recognition of the work done, that we were at once asked by the Ministry to undertake the urgent work of calculating sights for the Hotchkiss, Lewis and Vickers machine guns to meet the cases of air-planes flying with various speeds in various directions and at various low altitudes. The task was a novel and difficult one, but the theory was worked out in the Laboratory, and within four or five weeks of the order the tables were sent to France, arriving there just before Christmas. At present we are occupied with the wind influence on firing at high altitudes, and with new tables for naval high-angled guns, owing to the adoption of a new fuze by the Admiralty.

Samples of the war work of the Laboratory are enclosed in a portfolio accompanying this Report. It must of course be remembered that they are of a confidential character. They are evidence at any rate of the activity of the staff in my charge. I venture to think it would come as a grave blow to these young people to hear that at the present time the Court had not found it possible to maintain the grant. We have given of our best where it seemed from the national standpoint to be most urgently needed at the present moment, and it has meant work of a most strenuous and long-maintained character.

My object throughout has been to maintain a body of trained computers together who would have the force of character and the knowledge to meet new problems and remain as a nucleus for the Laboratory research work when peace returns.

(Signed) Karl Pearson.

A copy of the following letter from Vice-Admiral R. H. Bacon accompanied the Report:

Ministry of Munitions

Princes Street, Westminster, S.W. 1
14*th February*, 1918

Dear Professor Pearson,

Captain Moore has brought to my notice the letter which you addressed to him on the 13th February, and is, I understand, returning the communication from Major Douglas which accompanied it. I wish, however, to take this opportunity of expressing my cordial appreciation of the very valuable assistance which the laboratories under your charge have rendered to the Ministry in general, and to this Department in particular, and thereby to the Country during the War.

At a time of great pressure the Ministry found itself in need of a trained staff of computers provided with necessary machines to undertake gunnery work, and was indeed fortunate to find such a staff and machinery already in existence at the Drapers' Company Biometric Laboratory and the allied Galton Laboratory. The work done in connexion, *inter alia*, with the preparation of practically all the charts and high angle range tables for the a.a. guns of both Services has proved of inestimable value; and it was no small advantage to find at a time of national stress that a school had been trained in times of peace for such computing work as became, on the outbreak of war, a matter of such vital importance.

I leave it of course to your discretion to make what use you may please of this letter, but I think it would be very fitting if you were to bring to the notice of the Court of the Drapers' Company the very great value attached by the Ministry to the services which have been rendered by a laboratory which, I understand, owes much to the traditional generosity and public spirit of one of the great city companies.

Yours faithfully,
(Signed) R. H. BACON
Vice-Admiral, and Controller, Munitions Inventions.

APPENDIX IV

SUMMARY OF SUBJECTS DEALT WITH BY KARL PEARSON IN HIS TWO LECTURE COURSES ON THE THEORY OF STATISTICS GIVEN AT UNIVERSITY COLLEGE, LONDON, DURING THE SESSION 1921–1922 The material is taken from E.S.P.'s lecture notes of that date. (See pp. 208—211 above.)

FIRST YEAR COURSE, 1921–1922

First Term

Lecture*

1. Introductory outline; the conception of correlation as distinct from causation. Classification into qualitative categories, the contingency table and conception of independence. Characters on a quantitative scale, x and y; means of arrays, the correlation ratio; special case of the regression straight line; the coefficient of correlation.

5. Polynomial regression lines; the least square principle leading to the equating of moments. Definition and properties of moments; Sheppard's Corrections. Return to fitting polynomial regression lines; orthogonal functions. The study of variation about the regression line in an array leading to the study of frequency distributions.

9. Discrete variates: the binomial type of problem, cause groups at each trial independent; the hypergeometric type, cause groups not independent. Detailed study of the binomial; need for approximation leading to use of its moments. Attempt to graduate the binomial from the ratio of slope to ordinate leading to a differential equation. Solution of this equation gives (i) in special case the Normal curve, (ii) in general case, the Type III curve. Properties of the Type III curve.

* I have not indicated the point at which each individual lecture started but the figures in this column will show roughly how many lectures were given to different parts of the subject.

Lecture

14. Properties of the Normal curve. Representation of data on a normal scale. Gauss' work regarding the arithmetic mean, the mean square deviation and the method of least squares.

16. The Poisson limit to the binomial, its moments and uses.

17. Introduction of the more general differential equation whose solution gives Pearson's system of frequency curves. The symmetrical curves, Types II and VII. Types XII, V and IV. [End of term after 22nd lecture.]

Second Term

23. Completion of Type IV curve. Types I and VI. Special curves, VIII, IX, X and XI.

30. Recapitulation of work on frequency curves.

32. Corrections for grouping to be applied to moments. The Euler-Maclaurin Theorem. Abruptness corrections.

36. The fundamental problem of Statistics—to predict from the past what will happen in the future. Bayes' theorem, criticisms of Boole, Venn and others; the "equal distribution of ignorance." Suggested extension of this theorem. Work of Laplace, the Normal curve derived as an approximation to a hypergeometric series, the Type I curve as a much better approximation to this series.

42. "Probable error" theory. The sampling variation and co-variation in group frequencies. Large sample as distinct from small sample theory; original population assumed very large compared to sample, nature of approximation involved in substituting sample for unknown population values. The standard errors of moments calculated about a fixed origin, e.g. of the mean. [End of term after 43rd lecture.]

Third Term

44. Approximation to the standard error of a standard deviation. Use of R. A. Fisher's 3 and then n-dimensional space transformation to obtain sampling distributions of mean and standard deviation for a Normal population. Properties of the standard deviation distribution. Sampling moments of the squared standard deviation for any population.

48. Derivation of first-order approximations to the sampling moments and cross moments of moments. The standard errors of β_1 and β_2 to a first approximation; application of these results in choice of Pearson frequency type from observed β_1 and β_2.

51. The relation between variation in the means, standard deviations and frequencies of arrays in samples from a bivariate distribution; a test of significance for the squared correlation ratio.

55. The correction to be applied to the mean square contingency, ϕ^2, in a two-way table. [End of Session after 55th lecture.]

SECOND YEAR COURSE, 1921–1922

First Term

Lecture

1. Indices, approximations to their standard errors; spurious correlation.

2. Multiple correlation and regression; general problem: to predict x_0 from knowledge of $x_1, x_2, \ldots, x_n$. Regression function, $X_0 = \phi(x_1, \ldots, x_n)$; investigation confined to case where ϕ is a linear function of the x's. Constants in linear equation obtained by maximising correlation between x_0 and X_0; introduction of determinants in solution. Relation between n and possible magnitude of correlations of type r_{os} and r_{st} Application to prediction of characters in offspring from knowledge of characters in ancestors; the effect of environmental as compared with hereditary factors. Partial correlation.

6. Frequency surfaces for 2 or more variables; consequences of assumption that these depend upon a large number of underlying independent factors. The multivariate Normal surface. Special study of properties in the bivariate case; transformation to 2 independent variables, Sheppard's formula for calculating correlation.

10. The 4-fold table and tetrachoric coefficient of correlation; properties of tetrachoric functions.

13. The transformation to be applied to the mean square contingency, ϕ^2, to obtain an estimate of r, i.e. $C_2 = \sqrt{\phi^2/(1+\phi^2)}$. Class index corrections for broad grouping. Biserial and triserial r and η.

16. Standard errors of frequency constants in samples from bivariate distributions. $p_{uv} = \Sigma(n_{xy} x^u y^v)/N$. *Case* (*a*). x and y measured from fixed origin. Expectation of $(\delta p_{uv} \delta p_{u'v'})$; statistical differentials and mathematical differentials. Approximations used justified in case of large samples; no assumption of normality. Application to problems of selection Illustration in obtaining the standard error of r. Reference to R. A. Fisher's distribution of r and its great variety of frequency forms in case where population is Normal.

20. *Case* (*b*). Expectation of $(\delta p_{uv} \delta p_{u'v'})$ when x and y are measured from sample mean values; problems of selection again referred to; illustrations: $r_{\sigma_x . \sigma_y}$, $r_{r\,\sigma_x}$, $r_{r\,\bar{x}}$. [End of term after 21st lecture.]

Second Term

22. The Variate Difference Method of investigating correlation; discussion of some of Pearson's work in progress, effect of periodic terms.

28. Further applications of multiple correlation; the χ^2 test derived from consideration of correlated deviations in group frequencies. Problems associated with χ^2; the effect of substituting a fitted curve for population curve in goodness of fit tests.

31. Multiple regression formulae applied to inheritance; the Law of Ancestral Heredity, a *mass* law. Galton's work. General hypothesis of the Mendelian Theory; correlations to be expected to arise between relatives in a population mating at random. Form of the Ancestral Law to be expected on Mendelian theory.

Lecture

37. The standard error of the median. The standard error of the estimate of the standard deviation obtained from the quartiles; reliability of estimates of standard deviation calculated from other percentile points (all for a Normal population). Outline of method to be followed in determining the standard error of a "class index correction." [End of term after 39th lecture.]

Third Term

40. Further suggestions on the "class index correction" problem.
41. The method of Correlation of Ranks.
43. The standard error of the tetrachoric coefficient of correlation.
45. The analysis of compound material; the breaking up of a frequency distribution into two component Normal distributions.
48. Galton's Individual Difference Problem: the expected value of the difference between the pth and $(p+1)$th individuals in order of ranking, in a sample from a Normal population.
50. A further application of multiple correlation theory: the influence of selection applied to one or more characters on means, standard deviations and correlation coefficients. [End of Session after 50th lecture.]

BRIGHT'S DISEASE, NEPHRITIS AND ARTERIO-SCLEROSIS: A CONTRIBUTION TO THE HISTORY OF MEDICAL STATISTICS

By MAJOR GREENWOOD, F.R.S. AND W. T. RUSSELL

From the London School of Hygiene and Tropical Medicine

INTRODUCTION

IN the course of our ordinary academic and official work, we make daily, almost hourly, use of the various reports of the General Register Office. Increases or decreases of the rates of mortality from all causes and from particular causes naturally attract our attention. The vicissitudes of these rates arouse our curiosity, we wish to know how the changes recorded arise. We know that these arithmetical statements are the tabulations of opinions expressed by medical practitioners and that changes in them will be determined by three principal factors. The first is a change in the frequency of a cause of death the nature of which has always—which means, of course, for statistical purposes, during the last hundred years—been recognized by certifying practitioners and described by them in the same terms. The second is a change of opinion leading practitioners in one age to prefer cause A to cause B, but in another generation to prefer cause B, or perhaps to choose a cause C which their fathers had not recognized. The third, correlated with the second, is change in the grouping of causes, too numerous for separate tabulation, adopted by the central statistical department. All specific rates of mortality are influenced by the first and second factors, many by the third. The third factor cannot always be measured but can be ascertained by a diligent study of the official documents themselves. The second, on the other hand, usually presents an intellectual problem very difficult to solve. Nobody who is even slightly concerned with medical statistics is ignorant of the controversy regarding the real measure of increase of mortality from cancer. Almost all would agree that the decline of the rate of mortality from tuberculosis is one of the most certain of statistical results. Yet even in this matter, a reader of such a critical study as Rosenfeld's *Tuberkulosestatistik* (League of Nations, 1925, C.H. 284) must agree that there is room for doubt as to the extent of the change.

From the purely arithmetical aspect, mortality attributed to nephritis is a striking example of vicissitudes. Beginning as a statistically negligible factor it rose to be one of the arithmetically important causes of death, and rose steadily for decades. Then, shortly before the war, it began to decline and, although it

has never shown signs of reverting to its original statistical insignificance, is of much less importance than 25 years ago.

Although one of us has had no clinical training and neither of us any clinical experience more recent than 30 years ago, we were not ignorant of the fact that the work of the last century had greatly changed medical opinion respecting the importance of disease of the kidneys. It seemed clear enough that these changes of opinion must have affected the statistics greatly. On the other hand we knew vaguely that some forms of disease of the kidney had been attributed to intoxications which might arise in connexion with industrial processes not in use 60 years ago and we knew quite distinctly that one form of nephritis, that arising in connexion with scarlet fever, was much less frequent than a generation ago. Hence we could not be sure that the statistical changes were wholly changes of opinion. It seemed to us, therefore, that a general examination of nephritis as a statistical cause of death would be interesting. We were sensible of our lack of equipment for making this study, but we hoped that an imperfect attempt might stimulate better trained persons to enter on this field of research. Nephritis is only one of a number of problems awaiting examination.

Knowledge of nephritis

The ultimate data of the medical statistician are the opinions of medical practitioners and he is not, strictly speaking, concerned with the question whether those opinions are correct. So, a cynic might argue, we are not concerned with the scientific truth—even if we knew what it was—but with what ordinary men believed to be the truth. But, as the conclusions reached by scientific investigators do, eventually, become common property, it will be interesting before one concentrates upon common opinion to try to give a very brief account of the best opinion of successive generations.

The pre-Galenical physicians drew many and correct inferences from the character of a patient's urine, but they did not have, or are not known to have had, any clear ideas on the function of the kidneys. Galen was, of course, an accomplished physiologist. His account of the rôle of the kidneys in *de Usu Partium* (see Kühn's ed., vol. III, pp. 273, 362 *et seq.*) is clear and he had no doubt at all that the function of the kidneys was to separate waste matters from the whole of the blood; that was why, in his opinion, their veins and arteries were so large, and he mocked at those who held the urine to be a mere product of local metabolism, indeed his remarks on the reason for the dense texture of the kidneys might be taken for a foreshadowing of a filtration hypothesis. He had some knowledge of renal pathology and defined nephritis as a phlegmon of the kidneys accompanied with pain (Kühn, XIX, 424) He had an inkling of the relation between some forms of dropsy and visceral fibrosis and defined the vermicular pulse of patients with dropsy having this origin (Kühn, IX, 312). But Galen did not systematically co-ordinate his physiology,

pathology and clinical observations. In the thirteenth century William of Saliceto (see R. H. Major's *Classic Descriptions of Disease*, p. 483; Thomas, Springfield and Baltimore, 1932) recognized the association of dropsy with scanty urine in life, *durities in renibus* Ettmüller in the sixteenth century also recognized the connexion. But in the tractate on dropsy by the illustrious Sydenham there is no hint that he had even so clear an idea of the function of the excretory system as Galen. Dropsy he attributed to weakness (*debilitas*) of the blood. Sydenham's contemporaries Morton and Cole had more insight into pathology. But Morton's remarks on dropsy in phthisis deal with hepatic not renal functions (Morton, *Opera Medica*, ed. 1697, p. 131). It is a pity that Cole did not include renal functions in his general treatise *De secretione animali*. There are only cursory references (pp 135, 157).*

Sydenham was a great practitioner; he remarked in this tractate that just as Hippocrates blamed the officiousness of those who preferred speculation to practical observation, "so may a prudent physician of the present time blame those who believe that medicine is to be promoted by the new chemical inventions of our day, more than by any other process whatsoever" (sect. 23). He would, no doubt, have attached little importance to his contemporary Dekkers' (1648–1720) observation that in some urines a drop of acetic acid produces a white coagulum (Major, *op. cit.* p. 485) and not much to that of Cotugno (1736–1820) that the urine of a soldier with dropsy coagulated on heating. Cotugno's work appeared in 1765 (see Dock, *Annals of Medical History* (1922), IV, p. 287); nearly 60 years passed before anything was added. Then within two years of each other a paper by W. C. Wells and a treatise by John Blackall (abstracts in Major, *op. cit.* pp. 487–93) made Cotugno's point firmly. Both writers quoted several instances of dropsy with heat-coagulable urine, both referred to associated changes of the kidney of a sclerotic form in some cases followed to autopsy. Wells was an eminent London physician, Blackall a leading practitioner in Exeter. Richard Bright did not begin practice in London until three years after Wells' death but might have met him. There can be little doubt that these writings influenced Bright, who, so far as general opinion is concerned, is the pioneer of the subject.

Down to this year, Bright's, literally, epoch-making papers had not been reprinted in English, although a good abstract of the first and complete translation of the second formed volume XXV of *Klassiker der Medizin*, edited by the late Karl Sudhoff. Now an excellent edition, edited by Dr A. Arnold Osman, is available (*Original Papers of Richard Bright on Renal Disease*; Oxford Medical Publications, 1937).

Perhaps no single contribution to medical knowledge has received such early and general approbation as that of Bright. In Mackintosh's textbook (*Principles of Pathology and Practice of Physic*, by John Mackintosh, 4th ed., London, 1836)

* Printed in the edition of Morton's works above quoted.

Bright's first paper is abstracted at length, and for more than a generation the textbook accounts were simply abstracts of what Bright had said.

A reason for this rapid acceptance of Bright's teaching was its clarity and congruence with English habits of mind. In the first report (1827) Bright did not beat about the bush or indulge in speculations, but went straight to his mark. He recognized that many morbid conditions might be clinically associated with dropsy, but that when dropsy was associated with the excretion of albumen in the urine he had always found the kidneys to be diseased. He did not even claim that the disease of the kidney was necessarily the primary cause, but suggested that the altered excretory function might result from many factors either destroying the balance of the circulation or causing direct inflammatory changes in the kidney. It is material to note that of the 23 cases minutely described in this paper, many were of *acute* disease in young persons, indeed six of the patients were under 30 years and two under 20. In 9 of the 23 cases, there was a history of alcoholic intemperance.

Bright's next communication was in the Goulstonian lectures of 1833 (reprinted by Osman, pp. 153–66). Here again he recognizes the possibility of the renal disease being a secondary effect, but remarks (pp. 165–6) that he has observed the gradual approach and increase of cardiac hypertrophy coming on months after the albuminous condition of the urine had been established and suggests that cerebral symptoms may be due to the cardiac derangement, itself secondary to kidney disease.

Bright's report of 1836 is his fullest account and should be read carefully—by no means an irksome task for it is an admirable piece of writing. The title is significant: "Cases and Observations, illustrative of Renal Disease *accompanied with the Secretion of Albuminous Urine.*" The italics are, of course, ours. The point is that this criterion must increase the proportion of acute cases included in any sample of cases.

It would probably be correct to say that Bright regarded the passage of albumen into the urine as an essential element of the diagnosis, although he definitely says (pp. 96–7) that albumen may not always be found and is led to make some remarks (p. 98) which, had he been a Greek author or even Sydenham, would certainly have been claimed as an anticipation of views expressed after his time. Whether Bright would have regarded *any* morbid condition of the kidney associated with albuminuria as within his group is not certain. But, as Osman shows (p. 168), one of Bright's typical cases was of amyloid disease. We shall now briefly analyse the report.

Bright begins by expressing the opinion that this disease is amongst "the most frequent, as well as the most certain causes of death in some classes of the community, while it is of common occurrence in all; and I believe I speak within bounds, when I state, that not less than 500 die of it annually in London alone".

He gives next a vivid account of a typical history. At the outset we are conscious of a difficulty. The opening words of his typical history are these: "A child or an adult, is affected with scarlatina, or some other acute disease; or has indulged in the intemperate use of ardent spirits for a series of months or years." He then passes to the clinical description. A few sentences earlier he had spoken of intemperance having laid a foundation and remarked that "a more impressive warning against the intemperate use of ardent spirits cannot be derived from any other form of disease with which we are acquainted; since, most assuredly, by no other do so many individuals fall victims to this vice" (p. 94). But of the 10 cases described in detail the personal particulars are these. No. 1: The patient was a physician of 42 "who had always lived freely but not intemperately". He was first seen in 1832 and had had symptoms of albuminuria and dropsy for two years. He died of apoplexy in 1834. No. 2: The patient here was also a medical practitioner, aged 33, and he had had scarlet fever 9 years before Bright examined him. His symptoms went back several years and he died with signs of what we should call uraemia about 4 months after Bright had first seen him. No. 3: A youth of 17, who died in uraemia; he had had haematuria with calculus at 12. The fourth case was not of one of Bright's patients and the personal particulars are scanty. The fifth case was of a man of 25 who also died in convulsions and coma. He presented himself complaining of dyspepsia and dimness of sight. The sixth patient, a woman of 24 who died with convulsions and in an apoplexy, had a history of albuminous urine over 4 or 5 years. The seventh case, again of a young woman of 21, had a 15 months' history of dropsy. The eighth case was of an Irishman of 50 who died "with decided cerebral symptoms". The ninth of a marine aged 43 who had lived a very intemperate life. He died from peritonitis. The tenth was of a youth of 21 who also died from peritonitis.

Of these 10 patients, at least 6 were under 35 and one only was *certainly* an habitual spirit drinker. One case was evidently post-scarlatinal. In this particular sample, then, the aetiological importance of alcohol must have been slight. The 100 cases followed to autopsy recorded at the end of the memoir have few personal particulars except age which is recorded in 72.* These have the following age distribution:

– 5	0
–15	2
–25	7
–35	18
–45	16
–55	22
–65	6
–75	1
	72

* Bright (p. 151) says 74, but we can only find 72; probably a couple of figures were slipped in printing off.

So 37·5 per cent. of these were under 35, a much smaller proportion than in the sample of ten, but a considerable proportion.

We have already seen that the explanation of this is the criterion of selection, viz. by an albuminuria demonstrable with the means available a century ago. But, in his discussion, as distinct from his clinical records, Bright plainly attached much importance to alcoholism; he probably thought the non-acute cases of greater numerical importance than they had in his data. Thus he stressed the insidious nature of the disease and carried out some statistical experiments suggestive of Louis' methods. He had the urines of 130 patients in the wards of Guy's Hospital in the winter of 1828–9 tested for albumens. Of 130 tested, 18 had urine coagulable by heat and in 12 others traces of albumen were found. He then showed on a sample that the patients with coagulable urine had indeed disease of the kidney Friends and pupils at Guy's repeated the experiment on 300 and 141 patients, and Bright concluded that "the disease, in its various stages, from its earliest functional derangement to the confirmed organic malady, is one of the most frequent, as well as of most fatal occurrence: and I think I am fully borne out in the estimate, which I made at the commencement of this paper, that not less than 500 deaths annually occur in this metropolis, from this single disease" (p. 122). We should not be statisticians if we were not curious to learn how that particular figure was obtained. Bright does not satisfy our curiosity. However, those who like figures may be amused by the following.

In the year 1913, almost the *statistical* high-water mark of Bright's disease, 2009 deaths in London were classified under that heading. The population of the Administrative County of London in 1913 was perhaps two to three times that of Bright's London, so, *judged by this criterion*, his estimate was not excessive. In the memoir of 1836 Bright does not add materially to the pathological descriptions given and beautifully illustrated in his 1827 memoir. There he had reduced the macroscopic types to three. It is beyond our province to discuss these but it is fair to quote Bright's own words. "Although I hazard a conjecture as to the existence of these three different forms of disease, I am by no means confident of the correctness of this view. On the contrary it may be that the first form of *degeneracy* to which I refer never goes much beyond the first stage; and that all the other cases, including Sallaway, together with the second series, and the third, are to be considered only as modifications, and more or less advanced states of one and the same disease" (p. 70). Sallaway's case, as mentioned above, was one of amyloid disease.

Perhaps our account has been sufficiently detailed to justify the statement made earlier, that Bright's teaching was well calculated to produce an impression on the medical world. He had linked up such common afflictions as dropsy and such frequent modes of death as apoplexy and convulsions with a disease of the kidneys demonstrable by a simple test during life and upon a

plate after death. He had carefully refrained from exaggeration. He did *not* say, or even hint, that all cases of dropsy or all deaths from apoplexy or in convulsions were due to renal disease. He did not even suggest that an examination of the urine was an infallible diagnostic test. He *did* direct the attention of clinicians and pathologists to a new world of ideas. If any work deserves the epithet classical, his did.

Dieulafoy remarked many years ago but long after Bright's time that a signal merit of Bright was his caution, his abstention from dogmatizing about the aetiology of the morbid processes he described.

THE PATHOLOGY OF NEPHRITIS

To sketch the history of medical opinion even down to Bright's day was a little presumptuous in two statisticians whose combined stock of clinical experience is confined to what one of them learned as a medical student and assistant to a general practitioner 32 years ago. To go further is the height of rashness. But, if this memoir is to be of any use to the statistical reader, some attempt must be made to explain the difference between the point of view of a physician of Bright's professional standing now and that of Bright. In trying to give this explanation we shall probably fall into two snares; on the one hand we shall miss the significance of some points owing to technical ignorance; on the other something of what we say will be unintelligible to a statistician who has not read an elementary book on physiology. That is the inevitable fate of workers in borderland subjects. We ask for forgiveness.

Although Bright was a contemporary of the founders of modern pathology and physiology, the methods of research they began to use were not then instruments of precision. Bright's pathological anatomy was naked-eye anatomy; he could not have minutely investigated the detail of the changes in the kidney, or other organs, by means of serial sections and differential methods of staining the tissues. He had no means of exactly testing the biochemical efficiency of the kidney as an organ of excretion. The results of experimental interference with the kidney in other animals known to him were few. Even the technique of co-ordinating clinical records was in its infancy. He had to depend almost wholly upon his own observations.

The position Bright had reached was this. He had shown that a diseased state of the kidney might arise in a number of ways. First of all some poison, for instance that of scarlet fever, might excite an acute inflammatory condition of the essential elements of the kidney structure. The patient might be killed almost at once by this acute disease. That was one important set of cases. Then the patient to whom this accident had happened might survive, but survive maimed. Perhaps years afterwards, perhaps as the result of another accident, say some other acute illness, the *locus minoris resistentiae* gave way, and death resulted. That was another group, the acute became chronic and then either by

sudden failure or by slow deterioration, death resulted. Lastly, he had patients who had not sustained a pathological accident but through fault of constitution or fault of living (the alcoholic habit, perhaps) damaged their kidneys and eventually broke down with mortal symptoms and signs which could be debited to the renal function. Here there was never an acute stage of inflammatory disease at all. To use a favourite medical term, the onset was insidious. This group, not by any means a majority in his *statistical* experience, but important in his experience as a physician, has probably or certainly been the most important element in the official statistical record of Bright's disease and changes in knowledge of the aetiology of this kind of disease a principal factor of statistically recorded changes.

Progress since Bright's time has been of two kinds. On the one hand there has been what one might almost call a deductive process. Starting with the postulate that the kidney is diseased—we leave for a moment the question *how* diseased—one may set this problem. Why and how does the circulatory system become involved? Under what circumstances will dropsy, oedema, be produced? One has a problem in bio-hydrostatics, to the solution of which most of the great physiologists and pathologists of the last century have contributed. *Pari passu* one has an inductive or descriptive advance. An autopsy becomes not the examination of an hour or two but, in the aggregate, of months or years. The exhibits, as the writer of a detective story would say, in any one case give the micro-anatomy not of one but of many organs. Material is accumulated by means of which the micropathologist can deduce the chronological order of the morbid changes as between organ and organ and between one and another part of the same organ. Co-ordinated with this, one has an improving system of clinical records and therefore the means of correlating the stadia of anatomical change with symptomatic change.

Quite early in the history of advance, indeed less than 20 years after Bright's death, some pathologists had reached the conclusion that a proportion of the cases with, to use another favourite phrase, the clinical syndrome of Bright, took origin not in a primary lesion of the kidney but in a change in the kidney secondary to a degenerative process in the smaller arteries. In English textbooks, Gull and Sutton (1872) have the credit of pioneers. Whether justly or not, we have not inquired. At least out of such work developed the conception of arterio-sclerosis, which, as we shall see, became statistically of great importance in the early years of the century.

We do not propose to try to tell the story of the 60 years' progress. The two memoirs by Russell (1929) and Gray (1933)* will give any reader a view of the methods available. These authors on a foundation of micro-anatomical study, correlating their histopathology with clinical observations, reach conclusions.

* These memoirs form, respectively, Nos. 142 and 178 of the Medical Research Council's series of Special Reports.

These conclusions are not wholly concordant; it would have been strange indeed if they had been for the subject is very complex.

We cannot hope to give in a few sentences a correct view of the structures which have been the object of study; perhaps the following hints are not hopelessly misleading. Functionally the kidney consists of an immense number of tubules each of which follows an intricate course from a blunt end into which fits a vascular tuft, called the glomerulus, to junction with a main drainage conduit. The minute structure of the cells lining this tubule differs in different regions. The blood supply is extremely rich, some enters at the tuft, leaves it by a vessel of smaller calibre and then passes through many minute vessels; part of the tubule receives a blood supply which has not passed through the tuft at all.

The rich and peculiar disposition of the blood supply suggested nearly a century ago to the English physiologist Bowman a theory that urine was produced by a double mechanism, filtration of the inorganic constituents at the tuft, and separation of the organic constituents lower in the tubule, a theory which still, in greatly altered shape, commands approval. Clearly the working of this machine may be thrown out of gear by the destruction of any of its parts. If the blood supply is disordered, the machine cannot work, grist is not brought to the mill, and the millstones—the living cells of the secreting tubules—may crumble. Even with a normal circulation, however, the tubules might sustain damage, for instance by the conveyance to them in the blood of toxins. Finally, in the kidney as elsewhere, the habit of "nature" of not replacing valuable articles when broken in a packing case but of putting in more and more packing material and so often breaking what valuables are still there, has full scope; fibrosis, provoked by some injury, may bind up the machine, altogether distorting its fine structure.

Modern micropathologists have followed out in detail the changes arising in different parts of this system and correlated them with clinical states.

The two recent memoirs cited above bear out the last paragraph. Russell's is the more detailed on the histopathological side, Gray's on the clinical, but in both histopathological evidence is primary. The two investigators do not agree on all points, but their disagreement is not of much importance to us. The main difference of opinion is as to whether a circulatory disorder, an ischaemia, unaided by an inflammatory change, is capable of producing a serious or fatal functional disorder of the kidney as an organ of excretion. Russell thinks it is not: Gray dissents. Any attempt by an outsider to sum up the arguments would be impertinent. To such an outsider the division between inflammation defined as "the local reaction of living tissue against a damaging agent", the presence of which justifies the term nephritis, and its absence which requires the term nephrosis (see Russell, pp. 119–20) is fine. But, whatever may be the best *nomenclature*, the experts recognize the existence of two groups of cases which

Gray terms kidney of essential hypertension and arterio-sclerotic kidney, of quite different clinical significance and not primarily of renal origin. The former group, characterized anatomically by widespread change in the arteriolar system, in the smallest arteries, includes serious cases, which clinically may exhibit even mortal signs of renal insufficiency. The latter group generally do not show signs or symptoms immediately due to disease of the kidneys.

We have singled out these groups because both of them would probably have been included by Bright as examples of his disease; certainly the essential hypertension group, probably the arterio-sclerotic group. A certifier of our day would include the whole of the second group and some proportion of the first group under arterio-sclerosis. *What* proportion would depend upon the prominence of renal signs and the personal idiosyncracy of the certifier.

Summarizing this summary—perhaps we should, more modestly, follow Calverley and say curtailing the already curtailed cur—we infer that research has restricted appreciably the pathological connotation of Bright's disease.

Passing from the pathology to the aetiology, using that term in a general sense, we have not derived much information from either memoir, for the sufficient reason that general aetiology was hardly within the authors' terms of reference. Gray's investigation was based upon 500 consecutive autopsies and some proportional frequencies are available.

Under his classification, there were 7 cases of acute nephritis, 6 of subacute or early chronic nephritis, 10 of chronic nephritis, 43 of kidney of essential hypertension, 15 instances of severe arterio-sclerosis and no less than 357 with *some* evidence of arterio-sclerosis.

Leaving out of account the cases of acute nephritis (also those of acute nephrosis) the relative frequency of kidney of essential hypertension is seen to be so great as compared with that of chronic nephritis, that its attribution to this or that class by a certifier must be of immense statistical importance. But what, if any, relation there may be between the lesions of the various types and antecedent habits or constitution cannot be decided. We note that in the hypertension group the patients whose cases are recorded in detail, where a uraemic element was grafted upon essential hypertension, were aged 38, 31, 34, 49 and $8\frac{1}{2}$; only one then belonged to the middle period of life.

There may be somewhere a complete collection of the clinical histories of a large sample of patients. But then we should not have detailed pathological records of them because—a fact sometimes forgotten by statisticians—filling up schedules takes less time than cutting, staining and examining sections. Still one cannot escape a feeling that it *ought* to be possible to gain a little more knowledge of antecedents; actually we have opinions of very experienced physicians.

One of us many years ago extracted from the post-mortem books of the London Hospital the information recorded respecting persons aged 25 to 55 for the years 1889–1901. In order to form some idea of the change of nomenclature

the following experiment was done. We took out the first 300 and the last 300 of the records of males and noted the instances in which lesions of the kidney were recorded under cause of death. Taking the terms acute nephritis and parenchymatous nephritis, to mean an acute condition, granular kidney, nephritis, *chronic nephritis, interstitial nephritis, to mean a chronic condition*, the results were these:

In the first 300 (post-mortems of 1889–91) acute conditions were recorded in 2 and chronic in 25 instances. In the last 300 (1900–1), acute in 1, chronic in 20. Restricting the record to those in which the renal condition was the *only* entry, the first 300 had 0 acute and 8 chronic disease, the last 300 had 1 acute and 4 chronic disease. There is perhaps a faint suggestion of decreasing importance of renal conditions. Arterio-sclerosis appears in the record twice in the first series, the case of a man of 43 said to have had degeneration of kidney, and the case of a man of 51 with cardiac thrombosis. It does not occur at all in the last 300.

It seems clear enough that Bright and his contemporaries attached a good deal of importance to excesses of drinking and eating, to metabolic vices such as gout, to acute diseases such as scarlet fever and to poisons such as lead in the production of the diseases they were talking about. To those earlier statements we owe the crystallizing out of what may be called the popular conception of Bright's, *chronic* Bright's, disease. The patient is a middle-aged man who has worked hard, worried a good deal, and done himself well. But we need not continue, newspapers still print advertisements of quack medicines.

Of course the advertiser had a professional model usually of an earlier generation. One of us has sentimental reasons for thinking tenderly of a novel published nearly 40 years ago by a medical practitioner who had been a student in the 'seventies. Its hero, John Armstrong, who stifled remorse for an early misdeed by prodigies of work and skill as an operating surgeon, first notices that something has gone wrong with his vision. He consults an ophthalmologist who suddenly asks: "Have you any great mental worry?" John starts: "Do you think I have Bright's disease?" Next John becomes unconscious in the middle of a clinical lecture and eventually dies, whether by accidental drowning in another fit or by suicide is left an open question. John Armstrong's creator would have certified either chronic nephritis or chronic Bright's disease. Dr Gray would have certified essential hypertension; the son of John Armstrong's creator, either chronic nephritis or arterio-sclerosis.

The pathological interpretation of the hard working and high living more or less remorseful business man has shifted from the kidney to the arterial system. We hope rightly, although there have been sceptics.

At the beginning of the twentieth century (in 1904) one finds a clinician of the older school, Georges Dieulafoy who had been a pupil of Trousseau and was the author of an immensely popular textbook, growing sarcastic: "Tout cela est fort bien, et ces notions concernant l'artério-sclérose sont du plus grand intérêt, mais

il faut convenir néanmoins qu'on a depuis quelques années singulièrement abusé de l'artério-sclérose; elle est devenue envahissante, elle veut tout expliquer, et dès qu'on éprouve quelque difficulté sur telle ou telle interprétation pathogénique et même clinique, on vous répond: c'est l'artério-sclérose!" (Dieulafoy, *Manuel de Path. Intern.* 14th ed., vol. I, p. 854.)

A material factor upon which the older writers put stress was, as we have said, the abuse of food and drink, particularly drink, and the association between Bright's disease and intemperance was a commonplace (John Armstrong fortified himself with "stimulants"). It is usually illustrated by the specific mortality of the trades in which the use of alcohol is likely to be considerable. For instance in the occupational analysis of 1910–12; i.e. of the period when the *statistical* mortality of nephritis was near the maximum, in the standard population 33 deaths out of 790 were attributed to Bright's disease for the population of all occupied and retired males. Publicans and spirit and wine dealers had 79 in 1265, inn servants 52 in 1173, barmen 63 in 1724. Angina pectoris and arterio-sclerosis (grouped together) had 9 deaths for all occupied and retired, 15, 15 and 29 for the groups named.

Present views of aetiology would probably have increased the attribution to arterio-sclerosis (*vide infra*). Now take groups in which the toll of Bright's disease is still more excessive, although the occupations involve no peculiar temptations to or facilities for using alcohol, file-makers and cutlers; the former had 165 deaths from Bright's disease and 26 from angina pectoris and arterio-sclerosis, the latter 64 and 20. Here the work of Gye and Kettle has made it probable that one is dealing with a directly toxic effect of silicic acid upon the renal epithelium as well as a general change. Much more of this occupational nephritis is primary than of the nephritis of the drink trades.

We must then separate the general aetiological factors of the heterogeneous group into two classes.

Overstrain, excess of food and drink, worry, would be held to contribute to the arterio-sclerotic form.

Bacterial infections or specific toxic substances, such as the industrial poisons of silicon, some anilin products, chromium and lead, belong to the aetiology of the primary renal form.

Can we form a general opinion as to whether all or any of these general factors have varied? So far as the factors of arterio-sclerosis are concerned there is no uniformity of belief. That the industrial risks, increased in specific instances by the discovery of a new process or at specific times as during the war, have *generally* decreased is almost certain. That seems as far as one may fairly go.

When we come to the statistics, we shall have a little more to say on the question of aetiology, but it will not amount to much. We pass now from our obviously imperfect attempt to indicate the best opinion of the time to the easier task of abstracting what was taught to beginners in textbooks. The facts—to

use the actuarial term—of the statistician are the opinions not of Richard Bright or his representatives in successive generations but those opinions strained through the minds of all the medical teachers and modified, for better or worse, by the minds of the taught and their individual experiences. The textbook writer lags behind the genius and the average statistical practice of the certifiers in any generation will lag behind the teaching of the textbook then current, because many of the certifiers were taught by the previous generation. Again, in a living art no mere textbook can reproduce the spirit in which it is taught. Still something can be learned from a perusal of books once famous.

THE TEXTBOOKS

In the generation immediately following Bright's work, say between 1840 and 1870, we believe the most popular treatise on medicine was that of Sir Thomas Watson. "Few medical works have been more successful than this", wrote Munk, "it has passed through five large editions, and has enjoyed a greater popularity with students and practitioners than any similar book since the First Lines of Dr Cullen" (*Roll of the College of Physicians*, vol. III, p. 292). Munk was writing about 60 years ago. The third revised edition of Watson's *Lectures on the Principles and Practice of Physic* bears the date 1848. Diseases of the kidney are discussed in vol. II, pp. 563 *et seq.*, 614–55. Watson describes the naked-eye appearance of Bright's kidney and continues: "And what are the signs which indicate, to an instructed eye, the presence of these changes [he is speaking of the chronic form]? Some of them are precisely the same, in kind, as those which denote the acuter disorder; only mitigated in degree, and of slower march and succession. The patients are subject to obscure lumbar pains; to sickness from time to time, and to retching; and their urine is apt to be red, brown or dingy, as well as albuminous, from the intermixture of the colouring matter of the blood. They are obnoxious to inflammations of the serous membranes also: and more particularly to head affections, of which, often, they die; drowsiness, convulsions, apoplexy. And, to finish the resemblance, many of them, aye most of them, become at length anasarcous." There may be no albuminuria; coma he would regard with Christison as the natural termination. Cardiac disease is often present and may, he thinks, be secondary to the renal condition, but he doubts whether cardiac diasease can produce renal disease. He cites, without expressing full agreement, Christison's four rules for the differential diagnosis of renal from cardiac dropsy, viz.: (1) Most cases of febrile dropsy are renal. (2) When in anasarca the oedematous parts are elastic and do not pit on pressure—Watson is clearly sceptical of this. (3) When the dropsy is attended by diuresis. (4) When the specific gravity of the urine is less than 1010.

We think that the effect on the ordinary student of this teaching would be to encourage the certification of deaths where the symptoms were cerebral and the signs of dropsy present as due to renal disease; no suggestion is conveyed that the

renal symptomatology might be secondary to disease of the circulatory system. It is worth reminding a reader that of the total number of certificated deaths, only a small proportion would be certified after a post-mortem examination.

We do not know whether any other textbook in vogue between, say 1860 and 1880, had the popularity of Watson's treatise. A teacher of that epoch who wrote a textbook much admired by one of our own teachers (the late Dr F. J. Smith) was Charles Hilton Fagge. We have consulted the edition of 1886 (*Principles and Practice of Medicine*, by the late Charles Hilton Fagge; London, 1886). Fagge's discussion (vol. II, pp. 429 *et seq.*) is anatomically much more detailed than Watson's. He regarded Bright's disease as a "diffuse non-suppurative affection of the kidneys requiring to be regarded as a substantive disease from the clinical point of view". His clinical picture is less vivid than but not greatly different from Watson's. He observes that much the most important group of cases shows cardiac symptoms, 17 per cent. or more, and remarks that "until recently cases of heart failure secondary to cirrhosis of the kidneys were almost always regarded during life as examples of a primary morbus cordis".

Fagge's teaching would hardly stimulate any transfer from the cardiac to the renal group of certifications.

From 1890 to the present day we can trace the changes of teaching in successive editions of one famous textbook, that of Dr James Taylor. In the first edition of the *Manual of the Practice of Medicine* (1890) there were signs of a change of view, indicated by the following remarks:

"As to the nature of the thickening of the arteries, very different opinions have been expressed" (p. 688). "The cause of the cardiac hypertrophy in renal disease has been no less hotly debated than the many other conditions in this interesting disorder." "By some, chronic Bright's disease, in the form of granular kidney, is regarded as a general and simultaneous affection of the heart, the arteries and kidneys; but if this were true, we still have to account for the precisely similar changes which occur in chronic parenchymatous nephritis in which case the renal disorder undoubtedly precedes the other symptoms."

Under atheroma (p. 507) he observes: "But such diseased vessels frequently coexist with Bright's disease", and in the preface he says it is doubtful whether Bright's disease should not be regarded as a general disorder.

Taylor leaned in his teaching to the view of the previous generation but a doubt is suggested. Twenty-one years later in the 9th edition (1911), the passage quoted above, beginning "By some", is retained, but we find the remark (p. 909): "Much more characteristic is the more or less uniform thickening (arteriosclerosis) which affects the small arteries all over the body, as well as the vessels of the kidney itself." But (p. 647): "The subject of the symptoms of arteriosclerosis presents the difficulty that since the condition is almost certainly brought about in most instances by the action of toxins or poisons circulating in the blood, which produce excessive tension in the vessels, the toxaemia and the

increased tension are as likely to be responsible for the symptoms as the structural change in the arterial walls." Still a note of scepticism, suggestive of the passage quoted earlier from Dieulafoy, but arterio-sclerosis has arrived.

Passing on another 20 years one reaches the 14th edition (1930). We now have a complete distinction between the secondary contracted kidney—secondary here meaning that the change is secondary to an inflammatory process in the kidney itself—and the primary contracted kidney, primary because the process is part and parcel of a general arterio-sclerotic change. In the former one will find a dilute urine containing albumen, little tendency to oedema, the exitus letalis will (using the old-fashioned nomenclature) be with cerebral symptoms. In the latter the urine will be normal save for a trace of albumen, there will be dropsy and exitus letalis with cardiac or other circulatory symptoms.

The student of this edition has every reason to certify many deaths which 30 years before would have gone to nephritis as due to arterio-sclerosis.

It is an interesting speculation to consider how a student with a knowledge of the 14th edition of "Taylor" would classify those of Bright's cases recorded to have had hard or contracted kidneys Case 92, a man of 48 who died in an epileptic fit with convulsions whose heart is recorded to have been healthy, surely belongs to the secondary contracted group; case 24 of the woman aged 36 who died of anasarca presumably to the primary contracted group. One has the impression that a large proportion, perhaps a majority, would now have been classed to arterio-sclerosis and not to renal disease.

THE VIEWS OF MEDICAL STATISTICIANS

The reader is, we hope, now in the position of having *some* idea of the trend of scientific research and a rather clearer idea of the way in which knowledge was conveyed to the ordinary student of medicine. We pass now to the comments of the medical statisticians who annotated or drew inferences from the death certificates which came to the General Register Office. Of the successive commentators, Farr, Ogle, Tatham and Stevenson, Ogle alone could properly be said to have had much experience as a clinician and even his experience was limited to a comparatively few years during which he was assistant physician to a London teaching hospital. Farr had practised for a few years as a general practitioner; Tatham and Stevenson came to statistics from the Public Health Service. It is, on the whole, fair to say that, with the partial exception of Ogle, their knowledge of clinical medicine and pathology obtained after graduation was a book knowledge only.

Bright's conception of disease of the kidney already figured in the 1st Annual Report (see pp. 97, 105) and in the 2nd A.R. (p. 74) renal dropsy is mentioned. In the 4th A.R. a recommended nomenclature is printed, and renal dropsy and albuminuria are assigned to granular disease of the kidneys or *nephritis*, the term retained throughout Farr's period. In the 13th A.R. 430 deaths are assigned

to nephria and Farr remarks: "The apparent increase of the deaths by *Nephria* from 254 in 1847 to 430 in 1850 is partly referable to improvements in the diagnosis of the medical practitioners throughout the country: for the accuracy of these records depends as much on the diffusion of medical knowledge as on the progress of medical discovery. Nephria, as well as many diseases of the heart, were formerly referred to dropsy which is a common result of their protracted existence" (Appendix, p. 134).

In the 17th A.R. it is noted that the deaths assigned to nephria have reached 776 in 1854. In the 22nd A.R. (Appendix, p. 185) the increase, to 1258 in 1859, is thought to be due "to recent medical discoveries; some of the cases which were then (in 1850) classed under dropsy are now distinguished". In the 25th A.R. (Appendix, p. 189) we read: "some of these diseases nephritis and nephria (Bright's disease) increase largely; perhaps only in appearance, arising from a change due to the diffusion of pathological knowledge."

In both the 29th and 32nd A.R. the continued increases are noted. In the 32nd A.R. there is a reference to the "newly discovered cause of death *embolism*". In the 38th A.R. (p. 233) we read: "The diagnosis by chemical reagents, and as the result of pathological research, is more advanced now among not only the heads of the profession but among general practitioners. How much the increase of nephritis and the enormous increase of deaths referred to Bright's disease (nephria) is due to this cause it is difficult to decide." That concludes Farr's references to the matter.

He evidently attributed much of the increase to change of practice in certification. His successor, Ogle, had no doubt at all; on p. xvi of the Supplement to the 45th A.R. he writes: "There can be no reasonable doubt that the apparent increase of mortality from renal disease is attributable to the gradual extension of the knowledge of Dr Bright's discoveries and the recognition of cases as renal that previously were attributed to other causes. It is possible of course that there may also have been some real increase; but there is no evidence whatsoever that such has been the case." Ogle did not return to the subject, but in the 58th A.R. Tatham notes that the results of letters of inquiry when dropsy was the certified cause of death (a plan first begun in 1884) gave in 1885 attributions to the circulatory system and the kidneys in the ratio of 470 to 191, but in 1895 the ratio was 102 to 35. Farr had already noticed in the 'seventies that attributions of deaths to the circulatory system were increasing. Save for a passing reference in the 69th A.R. (p. cxiii), where the still continuing increase of mortality attributed to Bright's disease is attributed in part to improved certification, the question is not specifically raised again until the second decade of the twentieth century.

It is not, we think, unjust to say that Ogle was too sceptical of the value of statements on death certificates—the accuracy of which he took energetic measures to improve—to think attempts to evaluate increases or decreases of particular causes of death repaying, while Tatham was not much interested in

clinical statistics. Stevenson had the wide range of interests, the desire to find truth hidden behind figures, which characterized Farr so that after 1911 we enter on a new period of research. The first point to attract Stevenson's attention was the statistical importance of arterio-sclerosis, to which he devoted a critical study in the A.R. for 1916 (pp. lxxi–ii). He notes that in 1912 (the year following the first separate tabulation in our records) the Local Government Board gave much publicity to the subject of arterial disease, while in the same year the Registrar-General circulated suggestions to practitioners pointing out *inter alia* the inadequacy of "old age" as a certified cause of death. Deaths from arterio-sclerosis rose and the proportion per 1000 deaths at ages 75 assigned to old age fell from 346 in 1911 to 311 in 1912, 295 in 1913 and by 1916 to 267.

Although for arterio-sclerosis, so named, only six years' records were available, Dr Stevenson noted that a good indication of trend in certification was afforded by the heading "other diseases of blood vessels"; that heading included arterio-sclerosis of the present list and cerebral atheroma. The two made up in 1911 over 97 per cent. of the old heading Other Diseases of Blood Vessels. On this basis it seemed that the growth of mortality was from 38 per million in 1901 to 244 in 1916; the growth was particularly great after 1911.

Until 1901 nearly all deaths from diseases of cerebral arteries went to apoplexy. It was then customary to certify the condition resulting from cerebral haemorrhage; next a step back to the vascular lesion was taken; now a further step is made to the cause of the vascular lesion, arterio-sclerosis; a yet further step might be to pass from arterio-sclerosis to the cause of arterio-sclerosis, perhaps in some instances to Bright's disease and then again behind the kidney lesion to some toxin. In order to form some notion of the possibilities here, the deaths of males in London in 1914 assigned to arterio-sclerosis were tabulated by associated cause.

Of the 550, 218 mentioned only the arterial cause. Of the remainder, 47 recorded bronchitis and 29 chronic Bright's disease. Cerebral haemorrhage, apoplexy 84, other disease of heart 45. Cerebral haemorrhage then was losing more deaths than any other cause to the profit of arterio-sclerosis, although chronic Bright's disease lost 29, about 3 per cent. of the number actually assigned to chronic nephritis in London that year.

Ten years later (text volume of A.R. 1926, pp. 84 *et seq.*) Stevenson returned to the subject in an even more interesting essay. Now his principal object was to try to find a cause of death which might be an *index* of degenerative disease of the circulatory mechanism. He concluded that cerebral haemorrhage, taken in its widest sense, viz. to embrace thrombosis, embolism, etc., might serve. Unfortunately, he could not go back earlier than to 1901 because before then deaths certified to cerebral embolism or thrombosis went to the general heading embolism and thrombosis and could not be picked out. But basing himself on the quarter of a century's experience available, he came to the conclusion that the rate of

mortality was certainly not increasing, perhaps decreasing. He showed too that the increase of mortality from heart disease (confined to ages over 75 years) was fictitious, largely due to an increased habit of certifying myocardial degeneration in the aged.

He came decidedly to the conclusion that the increased mortality from both heart disease and from arterio-sclerosis was a matter of book-keeping and remarked: "So far, in fact, as the records of certification can show, alarmist pronouncements as to increase of mortality from heart disease by 'the stress and worry of modern life' may be met by the observation that it is declining at all periods of life."

Passing to the subject of chronic nephritis he quotes the standardized rates (*vide infra*) and remarks that: "from quite small dimensions this mortality had grown, presumably, to a large extent at least, as a result of increasing recognition, to a level in 1901–10 close to the maximum attained about 1914. This was succeeded by a very sudden fall, for males from 392 in 1915 to 288 in 1919, or 27 per cent. in four years, and for females from 287 in 1914 to 198 in 1919, or 31 per cent. in five years. Such a change in so short a period, and occurring between the dates mentioned, inevitably suggests the influence of war conditions, but if so this has been maintained since the peace, for since 1919 the rate for males has further fallen by 13 per cent. and that for females by 1 per cent." He points out a parallel fall in the mortalities attributed to alcoholism or to alcoholic nephritis. He suggests that: "the connexion between alcoholic excess and Bright's disease may be closer than might have been anticipated. Alcohol is commonly regarded both as a contributory cause of Bright's disease and as harmful to those who suffer from it. The latter fact may serve to explain why, if reduction of the supply of alcohol is assumed to account for the sudden fall in mortality from Bright's disease, this commenced as soon as the supply of alcohol was reduced, although its action in inducing the disease is presumably slow." He suggests that the continuance of the decline might be associated with the price of alcohol as one of the few wartime conditions still operative.

THE STATISTICS

We have now outlined the history of opinions held by certifiers and pass to the statistics compiled from their reports. Perhaps it will be said that, having regard to the history as related, manipulation of the data is idle. All they can give is a kind of numerical verification of changes of opinion. In Table I we have the statistically comparable figures for seven decennia. Here are included both acute and chronic Bright's disease. By 1891–1900 the rate has much more than doubled that of 1861–70, it rises a little more, then begins to fall, and by 1921–30 is appreciably less than it was 40 years earlier. The rise is entirely consistent with an increasing habit of certifying nephritis in preference to symptoms or signs —whether cerebral symptoms, coma, etc. or dropsy, etc.—and the fall is per-

fectly consistent with wider knowledge of ultimate aetiology leading to a re-transfer to arterio-sclerosis. Since the total rate of mortality from arterio-sclerosis is greater than that for nephritis, even allowing for an attribution to this group of a large number of cases which the older practitioners would certainly *not* have

TABLE I

Mean annual mortality per 1,000,000 *living. England and Wales.*
Acute and chronic nephritis

Periods	All ages standardized	0–	5–	10–	15–	20–	25–	35–	45–	55–	65–	75+
Males												
1861–70	153	71	52	34	44	66	118	206	303	434	584	647
1871–80	269	135	91	49	66	94	179	337	545	844	1128	1234
1881–90	364	184	95	55	78	113	195	404	718	1214	1869	2116
1891–1900	418	174	77	50	73	103	182	420	844	1556	2396	2831
1901–10	435	148	67	47	64	96	162	367	837	1750	2822	3415
1911–20	406	121	62	57	72	96	154	310	732	1548	2789	3724
1921–30	303	67	42	40	61	78	98	200	484	1064	2217	3887
Females												
1861–70	95	48	31	25	33	56	87	134	169	252	321	265
1871–80	179	97	58	44	62	93	146	240	324	501	643	629
1881–90	265	152	68	51	75	119	184	333	483	798	1180	1196
1891–1900	308	143	55	51	68	108	183	367	604	1012	1564	1691
1901–10	325	122	54	50	65	96	166	349	633	1163	1813	2095
1911–20	302	111	54	53	66	97	152	294	565	1036	1757	2296
1921–30	255	55	34	47	58	90	125	216	486	827	1582	2515
Increase or decrease per cent. compared with 1861–70												
Males												
1871–80	+ 76	+ 90	+ 75	+ 44	+ 50	+ 42	+ 52	+ 64	+ 80	+ 94	+ 93	+ 91
1881–90	+138	+159	+ 83	+ 62	+ 77	+ 71	+ 65	+ 96	+137	+180	+220	+227
1891–1900	+173	+145	+ 48	+ 47	+ 66	+ 56	+ 54	+104	+179	+259	+310	+338
1901–10	+184	+108	+ 29	+ 38	+ 45	+ 45	+ 37	+ 78	+176	+303	+383	+428
1911–20	+165	+ 70	+ 19	+ 68	+ 64	+ 45	+ 31	+ 50	+142	+257	+378	+476
1921–30	+ 98	− 6	− 20	+ 18	+ 39	+ 18	− 17	− 3	+ 60	+145	+271	+501
Females												
1871–80	+ 88	+102	+ 87	+ 76	+ 88	+ 66	+ 68	+ 79	+ 92	+ 99	+100	+137
1881–90	+179	+217	+119	+104	+127	+113	+111	+149	+186	+217	+268	+351
1891–1900	+224	+198	+ 77	+104	+106	+ 93	+110	+174	+257	+302	+387	+538
1901–10	+242	+154	+ 74	+100	+ 97	+ 71	+ 91	+160	+275	+362	+465	+691
1911–20	+218	+133	+ 74	+112	+103	+ 73	+ 75	+119	+238	+311	+447	+766
1921–30	+168	+ 15	+ 10	+ 88	+ 76	+ 61	+ 44	+ 61	+140	+228	+393	+849

certified as Bright's disease, one has here quite enough material to make the nephritis rate of 1921–30 as great as that of 1901–10. A transfer of about 25 per cent. of the total mortality assigned to arterio-sclerosis would suffice. This would, of course, be a very crude method. The age distribution of certified deaths by Bright's disease is similar to that of deaths by arterio-sclerosis to this extent, that both show increasing rates with increasing age. But the increase is very

much steeper in arterio-sclerosis. A statistically negligible number of deaths is certified to arterio-sclerosis at ages under 45 and in the age group 65–75 the rate of mortality is only a third of that at ages over 75. Bright's disease is statistically responsible for few deaths under 45 but at 65–75 its mortality rate is more than half that at 75– and at 55–65 as much as 23 per cent. of it. Any method of re-allocation would have to take account of this difference.

To those unfamiliar with the material, it would seem that it *ought* to be possible by the sorting out of items to construct rates of mortality comparable from generation to generation. Let it be agreed that the group of illnesses which Bright first described was a mixed bag; still they at least had this in common, that they were mortal illnesses. The certifiers of the deaths will have entered some striking feature, symptom or physical sign. Why not assemble from the deaths certified to dropsy, apoplexy, cardiac disease, arterial disease, etc., etc., those clinically concordant with Bright's set, form rates of mortality year by year and draw a conclusion? The answer is that the titular subdivisions of causes of death as published, or indeed as retained in unpublished form, are not sufficiently minute to permit of the reconstruction. We have quoted above illustrations, such as the difficulty Stevenson himself found in tracing the antecedents of arterio-sclerosis as a cause of death.

Having given much thought to the subject, we are obliged to confess that we see no way of statistically eliminating, by new rates, the changes of opinion through 60 years and, therefore, cannot measure the share of that change of opinion in the moulding of the conventional rates. A frontal attack on the statistics seems to us hopeless.

Regional data

There remains for consideration the question whether by taking the data in flank we can force them to tell us the truth. Our national data are tabulated in many subdivisions; the geographical, administrative and even occupational unit of tabulation are employed. Suppose we have a subdivision of the data into a large number of groups and suppose further that the standard or practice of certification among the medical attendants of those dying in the groups is uniform then, if we make the group rates of mortality from various causes the primary object of study, it should be possible by the statistical method to reach some interesting results.

Let us begin with the most obvious of groupings, into town and country or, virtually into areas of high and low density of population. This is, of course, not really a simple division at all; townsmen and countrymen differ in many things other than housing density. Still, let us leave it at that for the moment. Now suppose we take out a scheduled cause of death A, and find that its rate is higher in town than country (we assume that obvious sources of fallacy, age and sex distribution, transfer of deaths, etc., have been eliminated). If the practice of

certification is really the same in town and country then it is a fair conclusion that the disease is really in some aetiological connexion with town life (whether by procatarctic factors or by selection is an open question).

In Tables II (a) and II (b) we show the rates of mortality from nephritis for areas in descending order of urbanization at the epoch of recorded maximal incidence and at decennial intervals thereafter. Confining ourselves to the

TABLE II (a)

Death-rates per million in age groups, and the standardized death-rates from acute and chronic nephritis, according to degree of urbanization in England and Wales. (Males)

Periods	0–	5–	15–	25–	35–	45–	55–	65–	75+	Total	Standardized death-rate
London											
1911–14	55	41	70	151	388	1065	2264	3751	5354	528	528
1920–22	53	32	65	108	248	554	1259	2614	4418	394	345
1930–32	33	27	71	103	181	495	1097	2641	5662	437	336
County boroughs											
1911–14	151	72	92	189	442	1092	2187	3534	3912	514	533
1920–22	103	50	77	138	259	601	1352	2338	3555	381	351
1930–32	50	38	82	103	210	560	1230	2783	4870	443	354
Urban districts											
1911–14	132	56	79	145	335	834	1731	3149	3757	436	440
1920–22	79	44	59	120	231	497	1158	2172	3409	354	309
1930–32	51	32	60	91	165	478	1057	2486	4582	408	310
Rural districts											
1911–14	100	47	70	105	238	573	1246	2478	3178	394	332
1920–22	62	46	51	100	157	367	892	1816	3172	331	252
1930–32	54	29	61	71	156	382	905	2346	4660	424	284
Death-rates expressed as percentage of those in 1911–14											
London											
1911–14	100	100	100	100	100	100	100	100	100	100	100
1920–22	96	78	93	72	64	52	56	70	83	75	65
1930–32	60	66	101	68	47	46	48	70	106	83	64
County boroughs											
1911–14	100	100	100	100	100	100	100	100	100	100	100
1920–22	68	69	84	73	59	55	62	66	91	74	66
1930–32	33	53	89	54	48	51	56	79	124	86	66
Urban districts											
1911–14	100	100	100	100	100	100	100	100	100	100	100
1920–22	60	79	75	83	69	60	67	69	91	81	70
1930–32	39	57	76	63	49	57	61	79	122	94	70
Rural districts											
1911–14	100	100	100	100	100	100	100	100	100	100	100
1920–22	62	98	73	95	66	64	72	73	100	84	76
1930–32	54	62	87	68	66	67	73	95	147	108	86

arithmetically more reliable rates of later age groups, it is seen that in 1911–14 there was an immense difference between the rates of London and the County Boroughs on the one hand and those of the rural districts on the other. In the

TABLE II (*b*)

Death-rates per million in age groups, and the standardized death-rates from acute and chronic nephritis, according to degree of urbanization in England and Wales. (Females)

Periods	0–	5–	15–	25–	35–	45–	55–	65–	75+	Total	Standardized death-rate
London											
1911–14	60	48	67	123	375	847	1459	2344	3128	421	379
1920–22	37	47	61	91	187	415	946	1611	2473	305	244
1930–32	20	31	75	88	159	390	935	1951	3657	395	262
County boroughs											
1911–14	159	66	88	146	406	850	1435	2164	2618	410	391
1920–22	76	50	62	105	223	457	953	1585	2084	299	254
1930–32	41	38	75	96	201	465	965	2014	3399	392	279
Urban districts											
1911–14	106	51	70	124	310	617	1249	2008	2445	355	324
1920–22	56	46	51	102	199	404	810	1467	2154	287	229
1930–32	37	38	69	88	175	385	809	1694	3097	357	242
Rural districts											
1911–14	97	44	68	115	234	453	862	1650	2191	319	257
1920–22	73	39	55	106	182	352	646	1350	2104	289	210
1930–32	46	30	66	87	173	353	797	1737	3270	381	241
Death-rates expressed as percentage of those in 1911–14											
London											
1911–14	100	100	100	100	100	100	100	100	100	100	100
1920–22	62	98	91	74	50	49	65	69	79	72	64
1930–32	33	65	112	72	42	46	64	83	117	94	69
County boroughs											
1911–14	100	100	100	100	100	100	100	100	100	100	100
1920–22	48	76	70	72	55	54	66	73	80	73	65
1930–32	26	58	85	66	50	55	67	93	130	96	71
Urban districts											
1911–14	100	100	100	100	100	100	100	100	100	100	100
1920–22	53	90	73	82	64	65	65	73	88	81	71
1930–32	35	75	99	71	56	62	65	84	127	101	75
Rural districts											
1911–14	100	100	100	100	100	100	100	100	100	100	100
1920–22	75	89	81	92	78	78	75	82	96	91	82
1930–32	47	68	97	76	74	78	92	105	149	119	94

terms of Bright's original conception, this is what we should expect, another proof of

> O fortunatos nimium, sua si bona norint,
> Agricolas!

But it will be observed that at all important ages the town rates have declined more, usually much more, than the rural rates Among males 55–65 the London rate of 1930–2 is only 48 per cent. of its value in 1911–14, but the rate in rural districts was still 73 per cent. of its maximum. Perhaps this does mean what we should like it to mean, viz. that improvement was greatest, aetiologically, where most needed. The factor to which Stevenson alluded, viz. a decreasing abuse of alcohol, must have been more potent in town than country. But, our King Charles's head of certification practice is still with us. It was Stevenson's opinion, and one could have no better opinion, that the precision of certification was greater in towns, particularly in London, than in country districts. A difference correlated with the greater resort to hospital and institutions.

Perhaps the greater improvement of the town rates is no more than the result of town certification being more precise, or at any rate more sensitive to changes of medical opinion. If so, then we cannot tell how much of the excess of town over country is due to greater precision of nomenclature on certificates. We need hardly add that this criticism must be restrained within limits. It has never been suggested by the most sceptical that country practitioners or informants miss deaths altogether; it is certain that the death-rate from all causes together at later ages is lower in country than town.

There is still something to be urged upon the sceptic. The difference between the experiences of males and females is interesting. In London, between 1920–2 and 1930–2, the rates on males continued to fall at all ages under 65 (except 15–20); on females they were rising at 65–75 as well as at 75+. The standardized rate in females was higher in 1930–2 than in 1920–2, in males it was lower. In the county boroughs the increase extends back to the age group 45– in females. In the urban and rural districts there is less difference. It is very hard to believe that local certification practice is different as between the sexes. Of course it can be argued, and not unreasonably, that the general aetiological factors—intemperance, etc.—of the arterio-sclerotic group prevail more in the male sex. That would account for a smaller transfer from the "true" nephritis group to the arterio-sclerotic group, but hardly for an increase in nephritis. Here is something needing investigation.

One also notes something worthy of investigation in the trend of mortality in early life, although here we are mainly concerned with acute nephritis. There is an increased incidence in the age group 15–25. We cannot in the regional data trace this into finer age groupings, but for the whole country it is possible (Table III). It will be seen that only at 15–20 for males but at both 15–20 and 20–25 for females there have been increases between 1920–2 and 1930–2. Is this a consequence of the wartime environment on early life?

TABLE III

The death-rates per million in age groups from acute and chronic nephritis in England and Wales for the two triennial periods, 1920–2 *and* 1930–2

	Age groups											
	0–	5–	10–	15–	20–	25–	35–	45–	55–	65–	75+	All ages
Males												
1920–22	81	50	41	53	81	122	231	513	1169	2174	3473	364
1930–32	49	30	35	63	76	92	180	489	1085	2558	4793	427
Females												
1920–22	64	46	47	47	68	102	203	413	836	1492	2157	293
1930–32	39	28	45	64	78	91	181	406	867	1833	3290	378

Occupational grouping

There is another method of grouping not subject, or not so greatly subject, to some of the difficulties just mentioned; the grouping by occupations. It is, of course, true that some large occupational groups, for instance agricultural labourers, are wholly non-urban workers. But if one considers the large number of separate groups, we have not the same constant bias. There is no obvious reason why the certification practice of the medical men who attend tailors should differ from that of those who attend weavers.

But it is certain that the occupational classification adopted in the successive censuses has changed so greatly that the contents of even the main groups now formed are occupationally different from those of 10 or 20 years ago. We think, however, that this objection does not weigh much against the use of the groups formed for the statistical purpose of measuring the correlation of mortality rates. We accordingly performed the following statistical experiment. In the 1910–12 occupational analysis, Bright's disease is a scheduled heading and another is angina pectoris and arterio-sclerosis.

In 1921–3 we have as separate headings chronic nephritis and arterio-sclerosis. If our reading of medical opinion is correct, Bright's disease of 1910–12 will include a much larger proportion of deaths belonging to that group of Bright's own cases in which a general factor of pathological changes is involved than the chronic nephritis of 1921–3. Again the angina pectoris and arterio-sclerosis of 1910–12 will be much less representative of the general factor than the arterio-sclerosis of 1921–3.

If then we take as a third variable deaths from all other causes than the two named, we should expect that in 1910–12 Bright's disease would be more highly correlated and arterio-sclerosis less highly correlated with other mortality than in

1921–3. The experiment was tried, using in each case all occupational groups for which at least 10,000 years of life were available. For 1910–12 there were 115 and for 1921–3 there were 119 groups fulfilling this condition. The results are shown in Table IV.

TABLE IV

1910–12 (*taking industries with population over* 10,000)

	S.E.	N	Partial r S.E.
All other causes and Bright's disease	$r_{12} = \cdot 6407 \pm \cdot 055$	115	$r_{12\cdot 3} = \cdot 647 \pm \cdot 054$
All other causes and angina pectoris and arterio-sclerosis	$r_{13} = \cdot 0846 \pm \cdot 093$	115	$r_{13\cdot 2} = -\cdot 142 \pm \cdot 091$
Bright's disease and angina pectoris and arterio-sclerosis	$r_{23} = \cdot 2946 \pm \cdot 085$	115	$r_{23\cdot 1} = \cdot 314 \pm \cdot 084$

Subscripts: 1 = All other causes. 2 = Bright's disease. 3 = Angina pectoris and arterio-sclerosis.

1921–23 (*taking occupations with population over* 10,000)

All other causes and chronic nephritis	$r_{12} = \cdot 4743 \pm \cdot 071$	119	$r_{12\cdot 3} = \cdot 399 \pm \cdot 077$
All other causes and arterio-sclerosis	$r_{13} = \cdot 3984 \pm \cdot 077$	119	$r_{13\cdot 2} = \cdot 296 \pm \cdot 084$
Chronic nephritis and arterio-sclerosis	$r_{23} = \cdot 3191 \pm \cdot 082$	119	$r_{23\cdot 1} = \cdot 161 \pm \cdot 089$

Subscripts: 1 = All other causes. 2 = Chronic nephritis. 3 = Arterio-sclerosis.

Their interpretation deserves a little thought. If we have correctly interpreted the trend of scientific opinion, the aetiological factors of arterio-sclerosis and of that form of kidney disease which is secondary to an arterio-sclerosis are *general* factors making for deterioration, those of a genuine primary lesion of the kidney are more particular, due to nocive agents of specific character. If then certification perfectly reflected the current state of knowledge, we should expect to find considerable correlation between rates of mortality from the arterio-sclerotic group and from all other causes, and but little correlation between the rate from chronic nephritis and all other causes (we mean by all other causes the death-rate obtained after exclusion of nephritis and arterio-sclerosis). But when certification does not perfectly reflect such knowledge, there will be considerable correlation between the mortality from nephritis and from other causes and also considerable correlation between the mortality from arterio-sclerosis and from chronic nephritis. This will arise because in a group in which the arterio-sclerotic type of renal disease is prevalent, some practitioners will assign a death to arterio-sclerosis, others to chronic nephritis.

Now consider the arithmetical results. In the 1910–12 period, the partial correlations show a considerable association between Bright's disease and other causes, no significant association of arterio-sclerosis and other causes and a significant association of Bright's disease and arterio-sclerosis. In 1921–3, the

association of chronic nephritis with other causes is appreciably smaller, that of other causes with arterio-sclerosis appreciably larger and the association of chronic nephritis with arterio-sclerosis negligible.

We have in fact moved appreciably nearer to the anticipated position. It is unfortunate that the data of 1930–2 are not yet available for a further test. One can only speculate as to whether the chronic nephritis–other causes correlation will have declined still more and the arterio-sclerosis–other causes correlation increased again.

We do not, of course, expect that the chronic nephritis–other causes correlation will vanish. Even if all the really arterio-sclerotic kidney diseases pass to the arterio-sclerotic group, there must remain a link which will produce an arithmetical effect due to a genuine pathological common factor.

If the nephritis mortality rates in occupations are scrutinized in the light of our general knowledge, one sees that excesses fall into two groups. On the one hand, we have those groups in which an excessive use of alcohol is socially probable, the occupations concerned with the sale or manufacture of alcoholic drinks.

Actually the correlation between mortality from cirrhosis of the liver and

TABLE V

	Cirrhosis of liver		Chronic nephritis	
	C.M.F.	Ratio	C.M.F.	Ratio
Cellarmen	45·1	4698	66 6	1930
Brewers of ale, stout and porter	76·8	8000	55·9	1620
Inn-hotel keeper and publican	110·9	11,552	78·1	2264
Barmen	56·0	5833	88·7	2571
All occupied and retired	9·6	1000	34·5	1000

The C.M.F. (Comparative Mortality Figure) provides means of comparing the mortality at ages 20–65 experienced in different occupations, allowing for differences between the age distributions of the populations in those occupations. A standard population is chosen in such a way that the death rates at ages from all causes of all occupied and retired males would yield precisely 1000 deaths in that standard. The relative positions of other occupations are shown by the corresponding number of deaths reached when their death rates at ages from all causes are applied to the same standard population. The expected deaths are similarly found for different causes by applying the death rates at ages from the selected cause to the standard population. For example, the recorded death rates from cirrhosis of the liver experienced at different ages by all occupied and retired males produce 9·6 deaths in the standard population; the death rates of barmen produce 56·0 deaths. The total mortality of barmen is, therefore, 6 times the mortality of all occupied and retired males. This relative position is more clearly shown in the column headed ratio in which the C.M.F. of all occupied and retired males is taken as 1000 and those of other occupations expressed as ratios of that figure.

chronic nephritis can be brought out on the whole of the data. In the official report 162 occupations were used for correlation and the correlation of the comparative mortality figures of the two causes was shown by the Registrar-General to be $0{\cdot}419 \pm 0{\cdot}044$. Here we imagine one is dealing with the arterio-sclerotic form of nephritis. But one has a wholly different group of occupations in which the certified incidence of chronic nephritis is high.

TABLE VI

	Population	C.M.F.	Ratio
Rag grinders	3,556	77·2	2238
Pottery dippers and glaziers	2,117	79·5	2304
Coppersmiths	4,106	94·2	2730
Wool spinners and piercers	6,854	99·9	2986
Cotton and blow-room operatives	2,722	102·9	2983
Tin and copper miners (underground workers)	1,775	118·6	3438
File-cutters	1,425	215·1	6232
All occupied and retired	9,704,860	34·5	1000

The population is the number of workers between the ages of 20 and 65 years recorded in these occupations at the 1921 Census.

Here, for instance in pottery dippers and file cutters, we have occupations in which the toxic factors of lead and silicon are plainly of importance. In others, notably the textile groups, the aetiology is obscure. The two branches of textile workers have a mortality from nephritis three times greater than the average for all occupied and retired males. The Registrar-General in the Occupational Supplement for 1921–3 has described the position in the textile industry as follows: "the textile position is still more remarkable in regard to chronic nephritis, from which not one of the sixteen occupations fails to return mortality in excess of the average. With this nephritis excess, no doubt, is associated another distinctive mortality of textile workers—that from cerebral haemorrhage. From this cause only two of the sixteen textile occupations, wool sorters and wool weavers, fail to exceed the average mortality. It thus appears that the conditions of work in textile mills promote degenerative changes of the kidneys, heart, and blood vessels."

What are the incriminating factors responsible for these conditions? Does the type of shed, dry or humid, in which the cotton weavers work tend to promote these degenerative changes? The mortality occurring in moist and dry sheds, as given in the Occupational Mortality Supplement for 1921–3, is certainly suggestive, but, of course, one cannot assert that artificial humidity was the sole factor. The death-rates from cerebral haemorrhage and chronic nephritis in towns in which artificial humidity is used in the sheds are higher than those in which it is not used. On the other hand the mortality from circulatory disease is

considerably lower in wet sheds than in dry ones. The history of the mortality during 1921–3 was as follows:

TABLE VII

Disease	Cotton weavers in towns where artificial humidity in the majority of sheds			
	Was used		Was not used	
	C.M.F.	Ratio	C.M.F.	Ratio
Cerebral haemorrhage	62·1	1383	48·4	1078
Circulatory disease	110·5	726	245·8	1615
Chronic nephritis	45·0	1304	32 7	948
All causes	1065·0	1065	834·0	834

The ratio was obtained by relating the C.M.F. value for the particular disease to the C.M.F. value for the same disease amongst all occupied and retired males. Thus for Cerebral Haemorrhage the ratio was $\frac{62 \cdot 1}{44 \cdot 9}$ or 1383 and for All Causes $\frac{1065}{1000}$ or 1065.

One has here an interesting unsolved problem.

This completes our study of the statistical history of Bright's disease. We chose the subject partly because it is of interest in itself, partly as a characteristic example of a recorded rate of mortality in which it was *a priori* certain that changes of opinion had been an important factor of statistical variation. It seemed to us desirable that some pains should be devoted to a task unlikely to lead to brilliant discoveries. A topic of the same class, but much more hackneyed, is the share of changing opinion in the increase of mortality from cancer. We do not know that this subject has been investigated on the lines of our paper, viz. by bringing into relation with the statistics a history of opinion.

The practical conclusions to which we are led are these:

(1) We do not think that the statistical data of different generations can, by any practicable reclassification, be rendered sufficiently comparable to permit of any sound inferences as to whether the factors favouring "Bright's disease" did really increase or decrease through the last 60 years.

(2) We think that by 1921–3 a sufficient degree of uniformity in practice of certification had been reached to admit of statistical study of group mortality rates having some aetiological significance.

(3) A corollary of (2) is that the data of our own time will repay analysis by medical statisticians.

(4) A medico-statistical analysis of hospital data, bringing the clinical and personal particulars into still closer relation with the histopathological data, now so far more complete than in former times, would be of great value.

THE LONDON SKULL

By MATTHEW YOUNG, M.D., D.Sc.

1. *Introduction.* To Mr Warren R. Dawson belongs the merit of having recognized the importance of the fossilized fragment of a human cranium found during the excavations for Lloyd's new building in the city of London in 1925. On his advice the members of the Committee of Lloyd's submitted the specimen to Prof. (later Sir Grafton) Elliot Smith for scientific examination, and later very generously presented it to the Anatomical Museum of University College, London, where it is now readily accessible for inspection by anyone interested.

Mr Dawson has written the following account of the circumstances in which the fossil was discovered:

During 1924 and 1925 the site bounded by Leadenhall Street, Lime Street, and Leadenhall Place in the City of London, upon which the historic East India House originally stood, has been excavated for the erection of a new building for the Corporation of Lloyd's. Before the erection of the steel-work began, the central part of the site was cleared by means of mechanical excavators; but as the stanchions and girders rendered the available space more and more restricted, part of the digging had to be done by manual labour. The chance of finding fossil bones was slight in that part of the area in which the steam excavator was used, for this apparatus raises large masses of earth at each plunge and deposits its burden bodily into iron skips, which are in turn hoisted by cranes and emptied into lorries. On such parts of the site as were dug out by manual labour, however, fossil bones have from time to time come to light.

By the kind permission of the Committee of Lloyd's, facilities have been given for a scientific examination of these bones, and the clerk of the works, Mr G. T. Murton, has in all cases carefully noted the exact depths and the nature of the soil in which the finds were made. In March 1925 I exhibited three specimens at a meeting of the Zoological Society. These comprised the head of a femur and some molar teeth of the mammoth (*Elephas primigenius*), found in the river gravel at depths of 20 and 37 ft. respectively, and the ulna of a rhinoceros, which in Mr M. A. C. Hinton's opinion, may provisionally be referred to the species *R. antiquitatis* Fischer. This ulna came from the redeposited London clay, which at this spot underlies the gravel at a depth of 40 ft. It was actually found at a depth of 42 ft. These specimens have already been described and figured (Warren R. Dawson & M. A. C. Hinton, *Proc. Zool. Soc. Lond.* 1925, Part 2, p. 793). The rhinoceros bone has been presented to the British Museum by the Committee of Lloyd's.

At a later stage in the excavations some further remains came to light. Amongst these were the antlers and some limb-bones of the red deer (*Cervus elaphus*) from the river gravel at a depth of 30 ft., and the greater part of the skull of an ox from another part of the site at the 26 ft. level. The most interesting fossil, however, is part of a human skull, recovered from the blue clay, the same formation as that in which the remains of the woolly rhinoceros were found, and at exactly the same depth, 42 ft., but in the western portion of the site. The fragment of skull was broken into four pieces by a blow from the excavator's pick; and one of the pieces, a small triangular splinter from the anterior border of the parietal bone, was not recovered. The other three pieces were fitted together and exhibited at a meeting of the Zoological Society of London on 20th October. On that occasion the erroneous statement was

made that the skull was found at a depth of 26 ft from the surface; but a few days later the clerk of the works directed my attention to the error and informed me that the human fossil came from the blue clay in the 42 ft. level.

The misunderstanding arose from the fact that our inquiries concerning the "skull" were believed to refer to the remains of the ox found at a depth of 26 ft., and not to the flattened plates of bone, which were not recognized as parts of a skull.*

2. *Probable Age of the Skull.* The most probable age of the deposit in which the human remains were found has been the subject of much discussion. Mr M. A. C. Hinton,† of the British Museum, is convinced that it must be assigned to the third or lowest and latest of the terraces of the Thames, that generally known as the 20 ft. terrace containing the characteristic late Pleistocene fauna and the implements of later palaeolithic man—an Aurignacian horizon—but Mr C. N. Bromehead,‡ of the Geological Survey, is of the opinion that the fluviatile beds which underlie Leadenhall Street form part of the Taplow or Middle or 50 ft. terrace.

At Sir Grafton Elliot Smith's request Miss Dorothy A. E. Garrod examined the evidence for the age of the skull, and her final report which we are permitted to quote is as follows:

The opinion of Mr H. Dewey, of the Geological Survey, when I first consulted him about the age of the deposit was on the whole in favour of Mr Bromehead's opinion, since borings from the neighbourhood of the Lloyd's site showed the surface of the London clay, which forms the bedrock of the terrace, at an average height of 30–35 ft. O.D. Since that date, however, further work has been done on these deposits, and Mr Dewey has very kindly given me the result of his recent researches. Fresh investigations of borings in the immediate neighbourhood of Lloyd's have largely contradicted the older findings. Mr Dewey thinks the reason for this to be that many of the bores were started from basements, and in one particular case where this is now known to have happened it is found to give an error of 20 ft. in the O.D. height of the London clay. The general result of recent work has been to place the surface of the London clay in this area from 9–14 ft. O.D., and therefore to suggest that the deposits underlying Lloyd's do in fact belong to the Flood Plain Terrace, in spite of the fact that aggradation has brought their surface up to 50–55 ft. O.D. Mr Dewey tells me that he feels sufficiently sure of his ground to alter the mapping of the Leadenhall Street deposits, and to mark them as Flood Plain in the forthcoming survey of the London area. This being so, it remains to examine its bearing on the age of the London skull, which can no longer be considered as contemporary with the Taplow Terrace. Unfortunately, no implements were found in the redeposited clay of the Lloyd's section, but the fauna leaves no doubt that it is Pleistocene, and that it was laid down in a period of cold. In the Admiralty section, where the Flood Plain deposits were well exposed, Mr Lewis Abbott found Upper Palaeolithic implements in the uppermost part of the fluviatile beds By the courtesy of Mr Abbott I have been able to examine these, and I have no doubt that they are Aurignacian. Their presence in the upper part of the Flood Plain gravels does not, however, prove that the whole of the Flood Plain deposits are necessarily post-Mousterian. In the valley of the Somme, M. Commont found Levalloisian implements in gravels lying just

* This account is taken from Sir Grafton Elliot Smith's communication on the London skull in *Nature*, 1925, Vol. 116, p. 678.

† *Proc. Zool. Soc Lond.* 1925, Part 2, p. 793.

‡ *Nature*, 1925, Vol. 116, p. 819.

above river level, these gravels being capped by a deposit with Aurignacian tools. In the Thames Valley, however, no implements older than Aurignacian have so far actually been found in Flood Plain deposits, and it would be unwise to postulate for the Lloyd's skull an age more remote than the Upper Palaeolithic.

Mr K. P. Oakley, of the Department of Geology, British Museum, is of the opinion that Miss Garrod's report adequately sums up the position. He has very kindly given the writer permission to supplement this report by the following statement which he has prepared, and in which he summarizes the most recent views (June 1937) on the age of the stratum in which the skull was found:

A Taplow (Middle Levalloisian) age does not seem to be entirely ruled out by the revised level of the London clay bench, but even if the skull occurred in deposits belonging to the lower part of the Upper Flood Plain Terrace, there remains the possibility that it is of Late Levalloisian (Mousterian) age (see W. B. R. King & K. P. Oakley, "The Pleistocene succession in the lower parts of the Thames Valley", *Proc. prehist. Soc.* 1936, pp. 65–7), so that in conclusion one may say that the presumptive age of the skull must be at least Aurignacian, and very possibly older (Middle to Late Levalloisian).

Other authorities are not in agreement with Miss Garrod's view. In November 1925, Mr J. Reid Moir* expressed the opinion that the London skull was in all probability to be referred to the Mousterian epoch. In August 1932† he made the following further comments on the probable age of the skull: "The blue clay in which the specimen was found is well known and widespread over East Anglia. It is to be referred to the second inter-glacial epoch of that region, is usually surmounted by the upper chalky boulder clay, and is to be seen at High Lodge, Mildenhall, Hoxne and at Ipswich. In each case mentioned the clay contains unabraded examples of either late Acheulean or early Mousterian flint implements"

Sir Arthur Keith‡ is of the opinion that the evidence collected by geologists and by students of man's palaeolithic tools indicates a far greater antiquity for the London skull than that which has been assigned to it by Mr M. A. C. Hinton. He says: "the evidence is most definite that the gravel bed which covered the London skull was laid down before the Acheulean culture had reached its last phase and long before the Mousterian culture had begun", and that the skull should be considered to belong to the earlier and not to the later palaeolithic epoch.

While there is every reason for believing that the fragment of skull was naturally deposited and formed part of a human being who was a contemporary of the woolly rhinoceros, its age must be assumed to be at least Aurignacian, and very possibly older. In spite of differences of opinion on its exact age, the skull is undoubtedly that of the earliest Londoner yet discovered, and is thus of

* *The Times*, 4 November 1925.

† *Ibid.* 17 August 1932.

‡ *New Discoveries relating to the Antiquity of Man*, 1931, pp. 437 and 443.

sufficient interest to merit a detailed description of its characters and an inquiry into its affinities with other types.

3. *Brief Statements of Expressions of Opinion on the Skull by Sir Grafton Elliot Smith and Sir Arthur Keith.* The conclusions reached by the late Sir Grafton Elliot Smith from his examination of this skull are incorporated in two communications on the specimen to *Nature** and *The British Medical Journal*† in 1925, and in his essay on "The Human Brain"‡ published in 1927. In his view the skull reveals interesting primitive traits. The sagittal contour appears to be exceptionally flat and of distinctive outline. The endocranial cast, not only in the outline of its sagittal contour, but also in the modelling of its surface, shows some resemblance to the corresponding characters in the casts of female Neanderthal skulls from La Quina and Gibraltar, though the form of the cerebellum and the slightly greater fullness of the cerebral surface in the parietal region seems to indicate that the skull belonged to a rather primitive type of the species *sapiens* and was not Neanderthaloid. Sir Grafton also gave a detailed account of the evidence in support of his firmly held belief that the "Lady of Lloyd's" was left-handed.

Sir Arthur Keith§ is of the opinion that the London skull represents a modification of the human stock first revealed to us at Piltdown, Sussex. Its anatomical characters are those seen in the skulls of Piltdown and modern types of humanity, but its nearest affinity is to the Piltdown type.

4. *Anatomical and Morphological Features.* The principal anatomical and morphological features of the skull may now be described in some detail. The fragment includes the greater part of the occipital and left parietal bones and a portion of the right parietal. The figures in Plates I and II illustrate its general appearance from the lateral, vertical, occipital and internal aspects.

The portion of the occipital bone that is present comprises practically the entire squama occipitalis with the exception of a small part in the region of the left occipito-mastoid suture and a segment that appears to be relatively narrow, though this cannot be asserted positively, which formed the posterior boundary of the occipital foramen. The upper part of the squama to the extent of almost a third of the supra-inial arc is formed by a pre-interparietal bone, the transverse line of demarcation between which and the lower part of the squama is quite easily traced. The external occipital protuberance is fairly well defined but not very prominent, and the moderately arched superior curved lines which extend from the inion towards the lateral angles of the bone are very slightly elevated and rounded in their medial parts but fade away as they are traced laterally. The

* *Nature*, 1925, Vol. 116, p. 678.

† *Brit. Med. J.* 1925, Vol. 2, p. 854.

‡ *Essays on the Evolution of Man*, 2nd ed. 1927, p. 176.

§ *Op. cit.* pp. 463–467.

surface of the squama below the line presents relatively faintly developed muscular impressions. The right and left segments of the lambdoid suture that limit the bone laterally can be traced in their whole extent but are completely synostosed except in the lower fourth. The portions of the lambdoid suture bounding the pre-interparietal section of the bone laterally deviate appreciably from the curves of the lower and more extensive parts of the suture, so that the lambda lies farther forward than might be expected if the bone were of the normal type. On the endocranial aspect of the fragment, the line of junction of the occipital and parietal bones is clearly indicated at the lambda and as far down as the lower limit of the pre-interparietal segment of the bone by a definite groove, but it is not apparent below this level. The groove for the superior sagittal sinus is continued into the groove for the right lateral sinus as usually occurs. The level of the internal occipital protuberance is practically coincident with that of the inion, unlike the arrangement usually found in Neanderthaloid specimens. There is a definite asymmetry of the posterior cerebral fossae, the right being deeper than the left. This reversal of the normal asymmetry has been fully dealt with by Sir Grafton Elliot Smith in his discussion of the evidence of left-handedness.* The cerebellar fossae are exceptionally well defined and are separated by a prominent internal occipital crest. The crest terminates abruptly in a fractured surface, apparently at the point whereat it divides in most skulls into the two diverging ridges that limit laterally the small triangular area lying behind the occipital foramen. From a comparison of this feature with the corresponding features of complete skulls, it is possible to reconstruct tentatively the defective portion of the occipital bone, and so obtain the approximate position of the opisthion or midpoint of the posterior margin of the occipital foramen.

The left parietal bone is nearly complete but is deficient to some extent in its anterior area at the antero-superior and antero-inferior angles. A small wedge of bone has also been lost just below the midpoint of the anterior border. No part of the coronal suture is visible on the outer or inner surface of the most projecting parts of the anterior margin of the bone, but it may be assumed, with a fair measure of confidence, that the coronal line of fracture coincides approximately with the position of the suture from its relationship to the groove for the middle meningeal vein which is present on the endocranial aspect and is situated about 1 cm. behind the tip of the lower projection. This tip probably reaches to within 1 mm. of the coronal suture. On the sagittal border of the bone near its midpoint a short section of the sagittal suture is preserved. The lower border of the bone is intact in the great part of its extent. The section forming the parieto-mastoid suture line is practically complete. The portion articulating with the squamous part of the temporal is preserved in a sufficient part of its normal extent to give an indication that the parieto-squamous suture line was probably moderately well arched. The superior temporal line can be easily traced, though

* *Essays on the Evolution of Man*, 2nd ed. 1927, pp. 176–189.

not specially prominent, and appears to be situated relatively higher up on the parietal than in the average modern skull. The parietal eminence is not obtrusive. On the inner aspect of the bone the grooves for the anterior and posterior divisions of the meningeal vessels are clearly indicated. The thickness of the parietal bone on measurement at several points is found to be approximately 5 mm. on the average, i.e. much the same as the modern English female skull.

The portion of the right parietal bone that is preserved is a triangular area of which the base is formed by the lambdoid suture, and it comprises probably just under one-third of the original area of the bone.

The bones are of a reddish-brown colour and seem to be very heavily mineralized so far as can be judged from their weight, general appearance and the nature of the surface along the lines of fracture. The high degree of mineralization of the parietal bone is shown clearly in Plate III, where its radiograph is placed beside that of a parietal bone of much the same thickness belonging to a seventeenth-century Londoner's skull of the Whitechapel series. The condition of the cranial sutures (lambdoid and sagittal) suggests an age of more than 40, but probably less than 50 years. The smoothness of the contour of the skull, and especially the faintness of the muscular impressions in the occipital region, make it highly probable that the sex is female.

In its anatomical and morphological characters there is no clear evidence that the skull, though apparently relatively low and flat, should be considered of primitive type. The presence of the pre-interparietal bone has hitherto been considered a recently acquired feature which is relatively common in modern skulls but not seen in skulls of palaeolithic age. It may be noted, however, that a closely corresponding anomaly of the occipital bone is shown in the Saccopastore* skull, a female specimen belonging to the Neanderthal species, which was excavated near Rome and is of well-authenticated Mousterian age.

5. *Comparison of the London Skull with other Types.* In an attempt to throw further light on the affinities of the London skull, the measurements and relative proportions of the cranial characters that were determinable in the incomplete specimen have been compared in detail with the corresponding characters, where available, in the following skulls or groups of skulls: (*a*) a number of upper palaeolithic female skulls from European sites, (*b*) British skulls, including the Bury St Edmunds fragment, of reputed late palaeolithic or earlier date (excluding the Piltdown and Swanscombe specimens), (*c*) Seventeenth-century London and modern Scottish series, (*d*) female Neanderthal skulls from Gibraltar, La Quina and Saccopastore, (*e*) the Piltdown skull, and (*f*) the Swanscombe skull.

(*a*) *Comparison with the Solutrean and Aurignacian Female Skulls.* The comparison of the measurements and their relative proportions in the London fragment with the corresponding characters in the female upper palaeolithic

* Sergi, Sergio, *Rivista di Antropologia*, 1934, Vol. 30.

(chiefly Aurignacian) group is shown in Table I. It will be seen that in this relatively short series considerable variation is shown in the majority of the characters which are tabulated. Though no real emphasis can be laid on the mean values of the several characters as being stable or truly representative of the series, these have been tabulated for the characters measurable in the London skull and other important cranial characters which may be estimated approximately in this specimen. The characters in which the London skull deviates most notably from the corresponding characters in the Aurignacian series as a whole can be seen from the last column in Table I. On the assumption that the variabilities of the several characters in the Aurignacian group may be represented approximately by the standard deviations of the corresponding characters in the Scottish female series, the difference between the measurement of each character in the London skull and the mean value of the corresponding character in the Aurignacian group has been divided by the appropriate standard deviation. For the characters in which the ratio of the difference to the standard deviation is under 2·0 the London skull cannot be said to differ to a significant degree from the Aurignacian series. A survey of the figures in the last column shows that there are only six characters in which the ratio exceeds the value 2. These are S_2, S_2', $100 \times S_2'/S_2$, $100 \times$ biasterionic B/S_2', the height of the parietal arc and the parietal arc height index. All of these characters are associated with the length of the parietal bone or its curvature. Comment will be made later on the relatively short and flat parietal bone which, if our reconstruction at the bregma and the orientation of the skull can be relied upon, is a peculiar feature of the London skull.

Though we have taken the female skulls associated with the Aurignacian and Solutrean cultures together as a group, which is referred to, for brevity, as the Aurignacian series, it is more instructive to compare the London skull with the individual skulls in the group.

If our estimates of the maximum length (L) of the London skull as approximately 180 mm. and the basio-bregmatic height as relatively low, probably even as low as 120 mm., are accepted, while the maximum breadth (B) is known to be at least 144 mm. and not more than 146 mm., it is obvious that in size and proportions the London skull seems to differ very appreciably from many of the Aurignacian female skulls. The skulls which agree most closely with the London skull in length are No. V from Solutré (1924) and that from Obercassel. The Obercassel skull is, however, much narrower and probably much higher than the London skull, and the comparison with the London skull may generally be restricted to the Solutré skulls. Solutré I is appreciably longer and rather less wide but definitely higher than Solutré V, the basio-bregmatic heights being 132 and 123·5 mm. respectively. The Solutré V skull is undoubtedly relatively low and relatively wide in form. Its cephalic index is 81·0. Assuming that the estimate of the length of the London skull as 180 mm. is approximately accurate,

TABLE I

*Showing the comparison of the Principal Measurements of the London Skull with the corresponding Measurements in Female Upper Palaeolithic Skulls found in Europe**

Characters	The London skull	Cro Magnon No II	Solutré I (1923)	Solutré V (1924)	Grotte des Enfants (Cro Magnon type)	Grotte des Enfants (Negroid type)	Pred-most No IV	Pred-most No X	Ober-cassel	Lautsch II	Le Placard (1881)	La Vallée du Roc No I	Means	$\frac{\Delta}{S.D.}$ †
Maximum length (L)	—	192·5	185·5?	182·0	[195·0]	191·0	191·5	185·0	182·0	[182·0]	175·0	185·0	185·5 (9)	—
Maximum breadth (B)	144?	138·0?	146·0?	147·5	140·0??	130·5	143·5	140·5	129·5	132·0?	140·0	140·5	138·9 (11)	1·1
Basio-bregmatic height (H')	—	—	132·0?	123·5?	[120·5]	133·5	136·5	—	134·5	—	128·0	129·0??	131·0 (7)	—
Auricular height (OH)	—	[111·0]	115·0	[110·0]	—	114·5	[117·0]	106·5	111·5	[118·0]	—	—	111·9 (4)	—
Biasterionic breadth (B''')	114	—	115·5?	116·0??	—	—	—	104·0??	112·0	—	108·0	109·2?	110·8 (6)	0·8
Parietal arc (S_2)	105	131·0?	124·0??	110·5?	121·0?	134·5	132·5	115·5?	139·0	132·5?	118·0	133·0	126·5 (11)	−3·4
Occipital arc (S_3)	123	—	133·0??	130·0?	—	118·0	123·5	122·0	111·0	—	119·0	—	122·4 (7)	0·1
Lambda-inion arc ($S_{3\cdot1}$)	82	—	86·0??	85·5	[108·0]	71·0	73·5	70·0	63·0	—	—	54·0	71·9 (7)	1·3
Inion-opisthion arc ($S_{3\cdot2}$)	41	—	47·0	44·5	—	47·0	50·0	52·0	48·0	—	—	—	48·1 (6)	−1·4
Parietal chord ($S'2$)	101	123·0?	114·2?	102·9?	111·1?	120·9	119·7	105·7	125·0	117·5?	—	—	115·6 (9)	−3·1
Occipital chord ($S'3$)	100	—	104·2?	99·1?	—	98·4	100·6	96·4	94·6	—	—	—	98·9 (6)	0·2
Lambda-inion chord ($S'_{3\cdot1}$)	76	—	74·9	77·2?	[94·9]	66·4	68·1	65·9	59·7	—	—	—	68·7 (6)	1·1
Inion-opisthion chord ($S'_{3\cdot2}$)	40	—	46·1	42·9	—	45·1	48·9	50·3	46·1	—	—	—	46·6 (6)	−1·3
Height of parietal arc	12	23·0	19·0	18·0	21·0	26·0	24·0	20·0	27·0	—	—	—	22·3 (8)	−4·4
Height of lambda-inion arc	15	—	16·0	13·0	20·0	12·0	11·0	11·0	8·0	—	—	—	13·0 (7)	0·9
Lambda-inion-opisthion angle	117°	—	118°·0	108°·0	[110°·0]	120°·0	117°·0	112°·0	126°·0	—	—	—	116°·8 (6)	0·1
Length-breadth index (100 B/L)	—	71·7?	78·7?	81·0	—	68·3	74·9	75·9	71·2	[72·5]	80·0	75·9	75·3 (9)	—
Height-breadth index (100 B/H')	—	—	110·6?	119·4?	—	97·8?	105·1	—	96·3	—	109·4	108·9??	106·8 (7)	—
Index of parietal curvature (100 S'_2/S_2)	96·2	93·9?	92·1??	92·1??	91·8?	89·9	90·3	91·5?	89·9	88·7	—	—	91·1 (9)	2·9
Index of occipital curvature (100 S'_3/S_3)	81·3	—	78·3	76·2	—	83·4	81·5	79·0	85·2	—	—	—	80·6 (6)	0·4
Pearson's occipital index	58·1	—	56·0?	54·9?	—	60·1	58·2	56·4	62·3	—	—	—	58·0 (6)	0·1
Lambda-inion chord/arc	92·7	—	87·1	90·3	—	93·5	92·7	94·1	94·8	—	—	—	92·1 (6)	0·3‡
Inion-opisthion chord/arc	97·6	—	98·1	96·4	—	96·0	97·8	96·7	96·0	—	—	—	96·8 (6)	0·5‡
Lambda-inion arc/occipital arc	66·7	—	64·7	65·8	—	60·2	59·5	57·4	56·8	—	—	—	60·7 (6)	1·2
Inion-opisthion arc/lambda-inion arc	50·0	—	54·7	52·0	—	66·2	68·0	74·3	76·2	—	—	—	65·2 (6)	−0·9
Inion-opisthion chord/lambda-inion chord	52·6	—	61·5	55·6	—	67·9	71·8	76·3	77·2	—	—	—	68·4 (6)	—
Parietal arc height index	11·9	18·7	16·6	17·5	18·9	21·5	20·1	18·9	21·6	—	—	—	19·2 (8)	−4·4
Lambda-inion arc height index	19·7	—	21·4	16·8	—	18·1	16·2	16·7	13·4	—	—	—	17·1 (6)	1·0
Biasterionic B/maximum breadth	79·2	—	79·1	78·6	—	—	—	74·0	86·5	—	77·1	77·7	78·8 (6)	0·1
Biasterionic B/parietal chord	112·9	—	101·1	112·7	—	—	—	98·4	89·6	—	—	—	100·5 (4)	2·5
Biasterionic B/occipital chord	114·0	—	110·8	117·1	—	—	—	107·9	118·4	—	—	—	113·6 (4)	0·1

* The measurements of characters in the individual upper palaeolithic skulls are taken from the memoir by G. M. Morant in the *Annals of Eugenics*, 1930, Vol. 4

† *The standard deviations used here are those for a series of 47 modern Scottish female skulls given in Table II.* Δ *represents the deviation of the measurement* for the London skull from the upper palaeolithic mean

‡ See footnote § to Table II

with a breadth of 144 mm. the cephalic index of the specimen would be 80. If the basio-bregmatic height were as low as 116 mm.—which it might possibly be on the assumption that it holds the same proportion to cerebral height (i.e. the vertical distance from the subcerebral plane* to the vertex) as is shown in the Scottish series—the length-height index would be $100 \times 116/180 = 64{\cdot}4$. If the height were 120 mm. the length-height index would be 66·7, compared with 67·9 in the Solutré V skull. A basio-bregmatic height of 120 mm., or even 2 or 3 mm. greater, appears to be more probable than one less than 120 mm. The higher value is strongly supported from the relationships shown in the superimposition of the sagittal contour tracings of the London and Solutré V skulls (Fig. 1), to be referred to later. The London skull is thus probably very similar in its general form to the Solutré V skull.

On comparison of the other characters in the two skulls, we find that in biasterionic width the Solutré V skull is approximately equivalent to the London skull. Its parietal bone is almost as short as the estimated length in the London specimen, as is shown by the parietal chords of 102·9 and 101·0 mm. and the sagittal parietal arcs of 110·5 and 105·0 mm., respectively. The parietal curvature in Solutré V appears, however, to be rather, though not greatly, in excess of that estimated for the London skull, as shown by the respective ratios of chord to arc, 92·1 and 96·2. The lambda-opisthion (occipital) arc is even longer in the Solutré V and Solutré I skulls than in the London skull. Their upper segments (the lambda-inion arc) are relatively long as compared with their lower segments (the inion-opisthion arc). In this respect these two skulls resemble the London and differ strongly from all the other Aurignacian female skulls. The lambda-inion and inion-opisthion chords in Solutré V are of much the same size, both absolutely and relatively, as in the London skull. The corresponding characters in Solutré I are not greatly different from those in Solutré V, but in this respect it deviates more from the London skull than Solutré V does. The curvatures of the two segments of the occipital arc, as represented by the ratio of the chords to the arcs, in Solutré V are closely akin to those in the London skull. The ratio of the lambda-inion arc to the total occipital arc, and, as might be expected, the ratio of the two segments of the arc to one another, also agree fairly closely in the two skulls. Solutré V shows some divergence from the London skull in one character; its occipital bone as a whole is more curved. This divergence is brought out by an angle and two indices in the table. The lambda-inion-opisthion (occipital) angle is less obtuse in the Solutré V, being 108° as compared with 117° in the London, the index of occipital curvature ($100\,S_3'/S_3$) is 76·2 as compared with 81·3, and Pearson's occipital index in Solutré V is 54·9 as compared with 58·1 in the London skull. In regard to total occipital curvature Solutré I is nearer the London than is Solutré V. Three ratios

* For a detailed definition of the subcerebral plane, first used by Sir Arthur Keith, see the account given by him in *The Antiquity of Man*, 1916, p. 379.

are tabulated of which the biasterionic breadth is a component. The proportions which the biasterionic breadth shows to the maximum breadth and to the sagittal extent of the parietal chord in Solutré V are almost identical with the corresponding ratios in the London skull, while the ratio of the biasterionic breadth to the occipital chord (S'_3) is only slightly greater in the specimen from Solutré than in the London, the respective ratios being 117 and 114. So far as can be judged from the available measurements, the London skull would appear to be indistinguishable in type from the skull Solutré V.

The sagittal contour tracings of the London skull and the Solutré V skull are superimposed in Fig. 1. As neither the subcerebral plane nor the Frankfort plane was readily identifiable in both skulls (the first being determinable approximately in the London skull and the second alone indicated in the tracing of the Solutré skull) and since the bregma-inion chords were found to be identical, the method adopted was to make the two bregmas and the two inions coincident. The slight divergence in the two contours is brought out in this way, the London skull being fuller in the lower occipital and less full in the parietal region than the Solutré skull, though the divergence in the latter region may be to some extent accentuated by the deficiency that exists in the vault of the Aurignacian specimen from France.

(*b*) *Comparison with the British Skulls, including the Bury St Edmunds Fragment, of reputed Late Palaeolithic or Earlier Date, but excluding the Specimens found at Piltdown, Sussex, and Swanscombe, Kent.* Comparison of the London skull with upper palaeolithic female skulls has been mainly confined to those associated with this phase of culture found on the continent, as few well-authenticated specimens of the period have yet been found in England. A brief reference will be made here to the general features of the London fragment and to those in which the specimen resembles or differs from some skulls of late palaeolithic or earlier date found in England. Sir Arthur Keith gives a good summary of the general characters of these skulls.*

It must first be noted that if the estimated length of the London skull as 180 is approximately correct—the breadth being 144 or 146—then the cephalic index is 80 or 81, so that the skull must be considered of brachycephalic type although at the lower limit of the range. With a length of 180 mm. and a basio-bregmatic height probably about 120 mm., the altitudinal index would not be much greater than 67, so that the skull is definitely chamaecephalic.

Of the five human skulls which have been discovered in Aveline's Hole in the Mendips by the Spelaeological Society, University of Bristol, associated with an Azilio-Tardenoisian culture, two fall into the round category, their width being at least 80 per cent of their length, while three are relatively long. According to Keith, this is the earliest evidence of a brachycephalic people in England. All the skulls have, however, unlike the London specimen, a characteristically lofty

* *New Discoveries relating to the Antiquity of Man*, 1931.

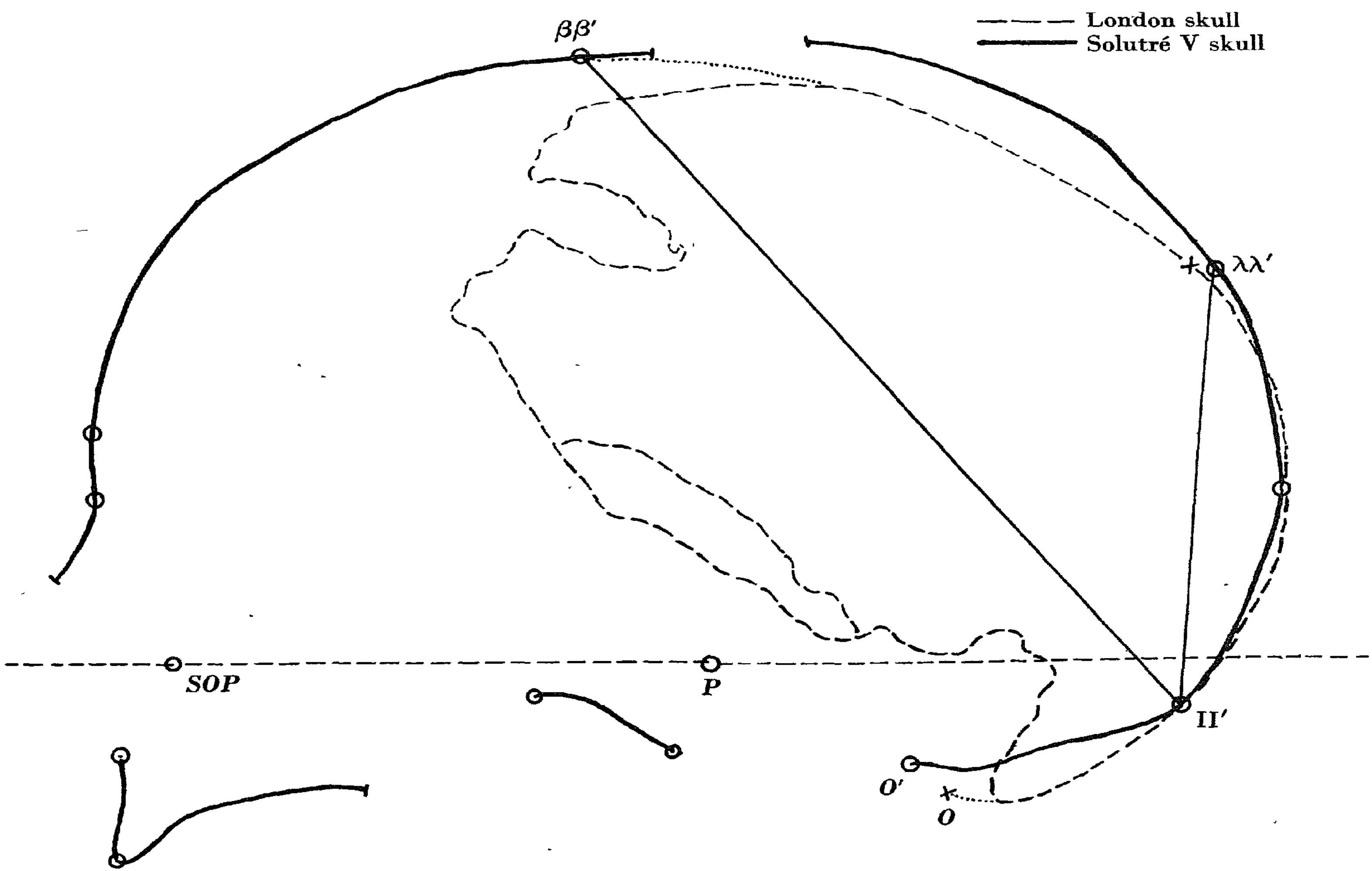


Fig. 1. Superimposition *of the median sagittal tracing of the London* skull on that of the Solutré V skull taken from Morant's memoir. The bregma-inion chord in the two skulls is identical and this chord has been made coincident.

vault. The famous Cheddar male skull which was found in Gough's cave in 1903 has a cubic capacity of 1450, definitely larger than that estimated for the London specimen. The length is 196 mm., the breadth 138 mm., the cephalic index 70·4, the height of the vault above the earholes 115 mm. and that above the subcerebral plane 105 mm. In all its characters the skull conforms to the river-bed type. In 1928 remains of five other skeletons were found in the same cave. Three of these were juvenile and one adult was represented only by a lower jaw. The remaining specimen was the calvaria of a young man, probably under 25 years of age, with an estimated cranial capacity of 1425 c.c., i.e. little different from that in the Cheddar No. 1 skull. With a length of 192 mm. and a breadth of 144 mm. the cephalic index of Cheddar No. 2 is 75, i.e. just within the long category and relatively broader than Cheddar No. 1. The vault is definitely lower, however, in Cheddar skull No. 2 than in No. 1, the respective subcerebral heights being 95 and 105 mm. The height of the vault of the skull Cheddar No. 2 is thus very little in excess of that of the London skull. All the skulls found so far at Cheddar are dolichocephalic.

Another specimen of special interest is that found in a fragmentary condition at Kent's Cavern, Torquay, in 1925. It is most probably of early post-glacial age. It is that of a young female, probably under 25 years of age. As reconstructed at the Museum of the Royal College of Surgeons, its estimated length is 175 mm. and maximum breadth 143 mm., giving a cephalic index of 81·7. The woman was thus definitely brachycephalic with an estimated cubic capacity of 1400 c.c. The skull had the same high vault as those found in the Aveline's Hole specimens, the highest point of the vault rising 120 mm. above the earholes. In this feature it differs characteristically from the London skull.

A brachycephalic skull is also said to have been found in the deposits at Cresswell Caves, Derbyshire, belonging to the same cultural stage as that shown at Aveline's Hole.

There is thus definite evidence of the presence of skulls in England in late palaeolithic times with length-breadth proportions of brachycephalic type, corresponding to that estimated for the London skull; but they are all definitely higher in the vault than it is.

Comparison of the London skull fragment with another cranial fragment found in England is of peculiar interest because, though the remarkable correspondence in certain of their features may merely be a coincidence, it is so striking as to suggest the possibility that it may be of special significance. The specimen in question is the Bury St Edmunds fragment, which can, according to Keith, be referred with a considerable degree of confidence to the later Acheulean phase of culture, immediately preceding the Mousterian phase to which, if not earlier, the London skull is believed possibly to belong by some recent investigators. Mr J. Reid Moir* is of the opinion that this skull, found in clay at Westley near

* *The Times*, 17 August 1932.

Bury St Edmunds and sometimes referred to as the Westley skull, is almost certainly of the same age as the London skull. An account of the fragment was published by H. Prigg* in 1885, but a more detailed account of its characters, accompanied by photographs of the different aspects oriented in relation to the subcerebral plane, was published by Keith† in 1912. The fossilized fragment of skull which is preserved in the Moyses Hall Museum, Bury St Edmunds, has been examined by the writer, but its comparison with the London skull fragment in this paper is mainly based on the photographs of the different aspects published by Keith in his memoir. The outlines of these were enlarged to life-size to facilitate comparison. The fragment consists of the upper two-thirds of the frontal bone and the anterior third of the right and left parietal bones. The upper region of the forehead presents a sharp "frontal bend". At the bend the frontal bone is comparatively thin and it is preserved intact sufficiently far forward to convince Keith of the practical impossibility that on such a forehead great simian eyebrow ridges were implanted. The characters of the specimen, according to Keith, clearly indicate a person with a head of the modern type, of the female sex (judging from the shape of the forehead and probable size), and probably over 40 years of age (judging from the condition of the sutures). Sir Arthur Keith made an attempt to reconstruct the probable outline of the missing parts of the skull and found its prototype in a skull, showing the same fronto-parietal contour, obtained from a gravel deposit in the East End of London and of uncertain antiquity. He used this as a guide in estimating the probable measurements of the missing parts. From the drawings of the lateral and upper aspects of the skull fragment when oriented in the subcerebral plane, Keith came to the following conclusions. After allowing for the missing parts of the frontal and parietal bones and the absent occipital bone, the length was probably 183 mm. It may have been shorter but not longer. The vault was remarkably flat, a character in which the Bury St Edmunds fragment resembles Neanderthal skulls. This flattening of the vault is probably natural and not due to soil pressure. Judging from the width and flattening of the vault the original transverse diameter of the skull could not have been less than 148 mm., the width being thus 80 or 81 per cent of the length. Such a skull would be classified as brachycephalic, but it is of a totally different type from most modern brachycephalic skulls, since the vault is so low. At the utmost the height of the vault above the ear holes could not have been more than 105 mm. The estimated brain capacity of such a skull, using the Lee-Pearson formula

$$(183 \times 148 \times 105 \times 0{\cdot}4 \times 206 = 1340 \text{ c.c.}),$$

is about equal to that of a modern Englishwoman. The Bury St Edmunds fragment, according to Keith, "is such a mutilated document that one may well

* *J. Anthrop. Inst.* 1885, Vol. 14, p. 51.
† *J. Anat. and Phys.* 1912, Vol. 47, p. 73.

hesitate in forming any certain conclusion as to the type of person it represents". A similar inference might reasonably be drawn from a study of the London skull fragment, and yet when the outlines of the lateral aspects of the two fragments are superimposed, making the bregmatic points and the coronal sutures coincident, it seems difficult to attribute what immediately becomes evident purely to chance. The superimpositions of the tracings of the lateral and upper aspects are given in Figs. 2 and 3.

Fig. 2 shows the Bury St Edmunds skull fragment, superimposed on the London skull fragment in such a way that the bregmatic points and the lines that indicate the sagittal contour tracing and the coronal suture tracing diverging from these points are all coincident. It will be recollected that the positions of the bregma and the coronal suture in the London fragment are only estimates, but they are estimations which can, from the extent of the parts preserved, be made with a fair degree of accuracy. When this superimposition is made it is found that the contour tracing of the Bury St Edmunds fragment is in almost exact agreement with the contour of the frontal region in the London skull which the writer had predicted as most probable from a knowledge of the measurements and their proportions in the London fragment. The most probable contour outline of the missing parietal and occipital region predicted by Sir Arthur Keith from the nature of the existing Bury St Edmunds fragment is rather less full in the occipital region than the contour as completed by the London fragment. Keith estimated the length of the parietal bone along its curvature as 126 mm., slightly greater than the frontal. The predicted position of the lambda is rather lower than in the London skull, but if allowance had been made for the fact that there is a definite inverse association between the lengths of the occipital and parietal bones, and that a long occipital arc that usually occurs when a pre-interparietal is present is likely to be accompanied by a relatively short parietal, the difference in the occipital region of the London skull and the predicted occipital outline of the Bury St Edmunds fragment by Keith would have been in still closer agreement. When the fragments are superimposed in the way described, the line of the subcerebral plane in the London skull—which can be estimated very approximately from the left asterion and the inion—exactly coincides with the subcerebral plane which Sir Arthur Keith had inserted in his reconstructed drawing for the Bury St Edmunds fragment. The subcerebral plane passes through the posterior inferior angle of the parietal and as a rule through, or just above, the fronto-malar junction. When a skull is so placed the highest point of the vault is situated in the majority of cases about 8 mm. above the level of the bregma and about 40 mm. behind it. It was on this principle that Keith oriented the Bury St Edmunds fragment. In any manner of orientation and measurement the vault of the Bury St Edmunds skull is a low one: the highest point of the vault is only 91 mm. above the subcerebral plane. This estimate coincides almost exactly with the corresponding estimate of the height

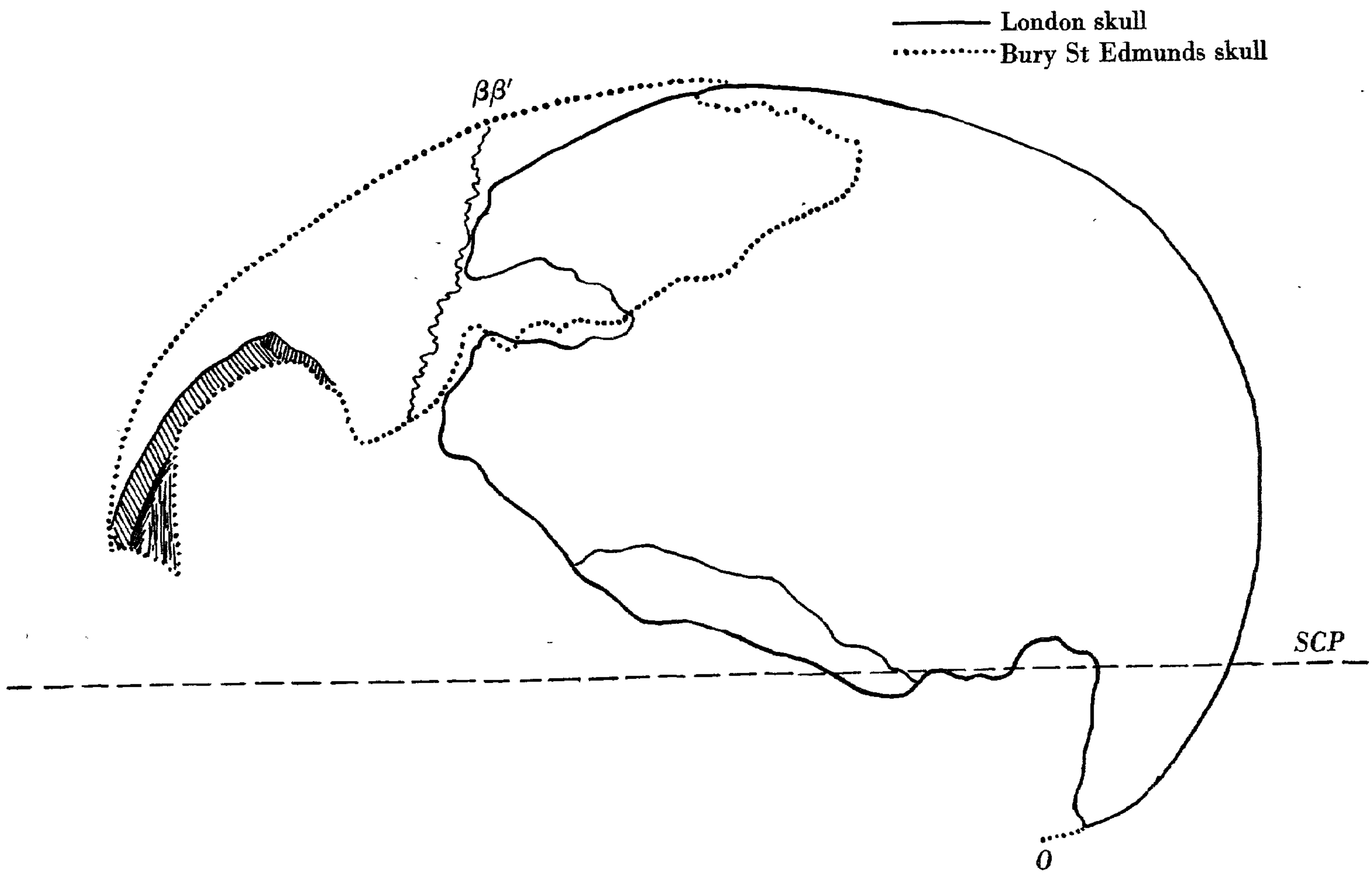


Fig. 2. Superimposition of median sagittal tracings of the London and Bury St Edmunds fragments with the bregmas and the adjacent corresponding parts of the coronal and sagittal sutures made coincident.

of the vault of the London fragment above the same plane. The slope or angle of the coronal suture is of assistance in verifying the plane of orientation and the angle is set as in the modern type.

Keith gives a view of the Bury St Edmunds fragment from above when oriented in the subcerebral plane, with an approximate outline of the complete skull indicated. The projected outline was largely determined from a comparison of the fragment with crania showing the same characters. This writer estimates that the maximum width of the skull could not have been less than 148 mm., and the proportion of width to the length is thus estimated to have been about 80 or 81 per cent., which is practically the same as our estimate of the cephalic index in the London fragment. Keith estimates the interstephanic diameter of the Bury St Edmunds fragment to have been about 110 mm. (the stephanion being the point where the temporal line for the attachment of the temporal muscle crosses the coronal suture). The corresponding measurement on the London skull fragment can be estimated approximately by doubling the measurement on the left side, and it seems to be of much the same order as the estimate for the Bury St Edmunds fragment. Fig. 3 shows the Bury St Edmunds fragment viewed from above when oriented in the subcerebral plane and superimposed on the London skull fragment oriented in the same plane; the bregmatic points as well as the sagittal and coronal sutural lines in the two tracings being made coincident. The close approximation of the London fragment contour to the contour completed by Keith for the Bury St Edmunds fragment is very striking; in the two specimens the estimated lengths are almost the same; the estimated maximum breadths and the estimated interstephanic breadths are also approximately the same.

We must fully acknowledge the risk of inaccuracy that is entailed in attempting to predict, even approximately, from such relatively small fragments as are preserved of the London and Bury St Edmunds specimens the characters and dimensions of the missing parts, yet in either case the predictions of the missing parts in one fragment are in remarkably close agreement with what is extant in the other. While this similarity may merely be a coincidence, it seems rather, in our view, to support the thesis that the two fragments represent calvariae that were essentially similar in type, being practically of the same length, about the same width and consequently showing an almost identical proportion of breadth to length, or cephalic index, which is at the lower limit of the brachycephalic range. So far as can be judged from the remains, both cranial vaults showed the same flattened contour and were equally low in their relation to the subcerebral plane.

(c) *Comparison of the Characters of the London Skull with those of* (1) *a Racially Homogeneous Group of Female Seventeenth-Century Londoners*, (2) *a Relatively Homogeneous Sample of forty-seven Modern Scottish Female Skulls without Interparietals, of a Type closely akin to that of the London Group, and* (3) *a Group*

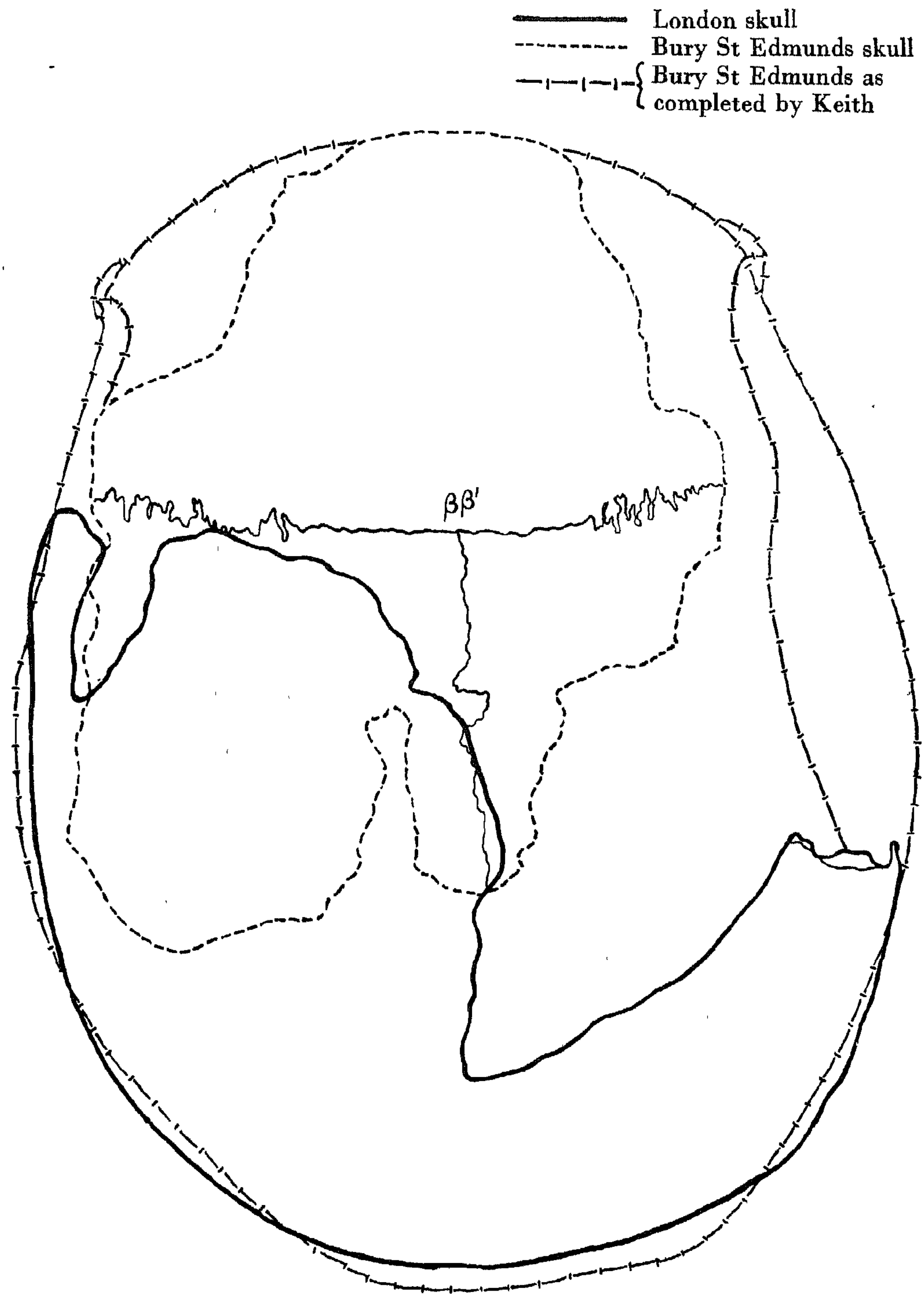


Fig. 3. Superimposition of tracings of the London and Bury St Edmunds fragments seen from the upper aspect with the bregmas and the adjacent corresponding parts of the coronal and sagittal sutures made coincident. The skulls are oriented in the subcerebral plane.

of ten Female Skulls with Interparietals from the same Scottish Series. The several features in which the London skull resembles and those in which it differs from the female skulls of Aurignacian and Solutrean date found in Europe and the female skulls relating to the upper palaeolithic phase in England have already been indicated and discussed. As the cranial fragment in some of its features appeared to approximate to the modern cranial type, it seemed not improbable that some further light might be thrown on its affinities by a comparison of its characters, in so far as they could be determined by measurement, with those of a relatively homogeneous series of modern skulls. This procedure will at least permit due consideration being given to the degree of variation that is normally shown in the several cranial characters in such a series. Whenever possible a comparison of this nature seems to be very desirable, if not essential, before appraising finally the status of a solitary, and possibly incomplete, skull that may be discovered, because of the definite tendency evinced by craniologists to assume that such a find may be considered a more or less average specimen of the population it represents. While, as a rule, the odds are greatly in favour of such an assumption being true, the possibility that the new discovery may be a normal, yet rather extreme, variant and not a truly representative or average specimen of a particular cranial type should always be borne in mind.

For comparison with the London skull, a suitable series of modern female skulls is fortunately available. This series forms part of the large collection of modern skulls, exceeding a thousand in number, which is preserved in the Anatomical museums at the University and in St Mungo's College, Glasgow, fully 700 being at the University and 300 at the College. Of the total series, approximately 400 are considered to be of the female sex. These female skulls exhibit a very close resemblance to those of the same sex in the collection of seventeenth-century Londoners from Whitechapel, the characters of which were described by Macdonell in 1904 in Vol. 3 of *Biometrika*. The female skulls in the Farringdon Street collection of contemporary Londoners described by Miss Hooke in 1926 in Vol. 18 of *Biometrika* are closely similar in type to those found in Whitechapel. The rather striking degree of similarity in the cranial series from Glasgow and London is described and discussed in a paper by the writer entitled "The West Scottish Skull and its Affinities".* Prof. Bryce readily granted permission for some further observations to be made on the female skulls included in the collection.

Reference has already been made to the fact that in the London skull the occipital bone presents the peculiarity that its supra-occipital segment had been divided at one time into two distinct parts, an upper and a lower, by a transverse suture which had become synostosed. This suture was situated some distance above the biasterionic diameter, so, in other words, a pre-interparietal bone had

* *Biometrika*, 1931, Vol. 23, pp. 10–22.

been present. Emphasis is laid on the presence of this anomaly in the skull, as it would appear to have a definite influence in determining the sagittal extent and also the curvature of the occipital region.

In Fig. 4 is shown the superimposition of the median sagittal contour tracing of the London skull on the corresponding contours of two relatively low-vaulted members of the Scottish female series, one (S 73) with, the other (C 45) without, an interparietal bone. The skulls are oriented in the subcerebral plane, the asterions being made coincident. The contour of the skull with the interparietal present seems to be in somewhat closer agreement with the London contour than the other.

Some years ago, during the systematic examination and measurement of the arcs and chords of the bones of the cranial vault in the long series of Scottish skulls, it was noticed that the presence of an interparietal in a skull was usually associated with an occipital arc and an occipital chord which were definitely above the average dimensions, just as an increase in average minimum frontal breadth characterized the skulls in which there was a persistence of the metopic suture.

As it was known that in the Scottish series of females there were included ten or twelve skulls in which an interparietal or a pre-interparietal bone was or had been present, it seemed to be advisable to compare the available measurements of the London skull with the average values for the "interparietal" group, as well as with the averages for a larger series in which there was no evidence that an interparietal had ever been present. The female skulls that exhibited a pre-interparietal or an interparietal, the serial numbers of which were known, were thus first extracted from the collection and set apart as one group. From the remainder of the series, comprising nearly 400 skulls, a random group of about fifty skulls was segregated by extracting specimens at more or less regular intervals throughout its extent.

After orientation in the Frankfort plane, accurate tracings were made by the dioptograph of the left lateral aspect of each of the skulls, including the profile or median contour from the bregma to the opisthion. On these tracings certain points and sutures were indicated, viz. the lambda, the inion, the left asterion, the left porion, and the left suborbital point, as well as the fronto-malar, the coronal, the spheno-parietal, the squamous and the parieto-mastoid sutures. The porion with the suborbital point, and the asterion with the point at the outer end of the sutural junction of the external angular process of the frontal bone with the malar bone provide the data necessary for the insertion of the lines representing the Frankfort plane and the subcerebral plane of Keith, respectively. The various chords were then drawn on the outlines and several measurements were taken corresponding to the measurements that were determinable on the projection of the median sagittal contour of the London skull fragment. The arcs and chords of the total parietal and occipital sagittal sections, together with those of the

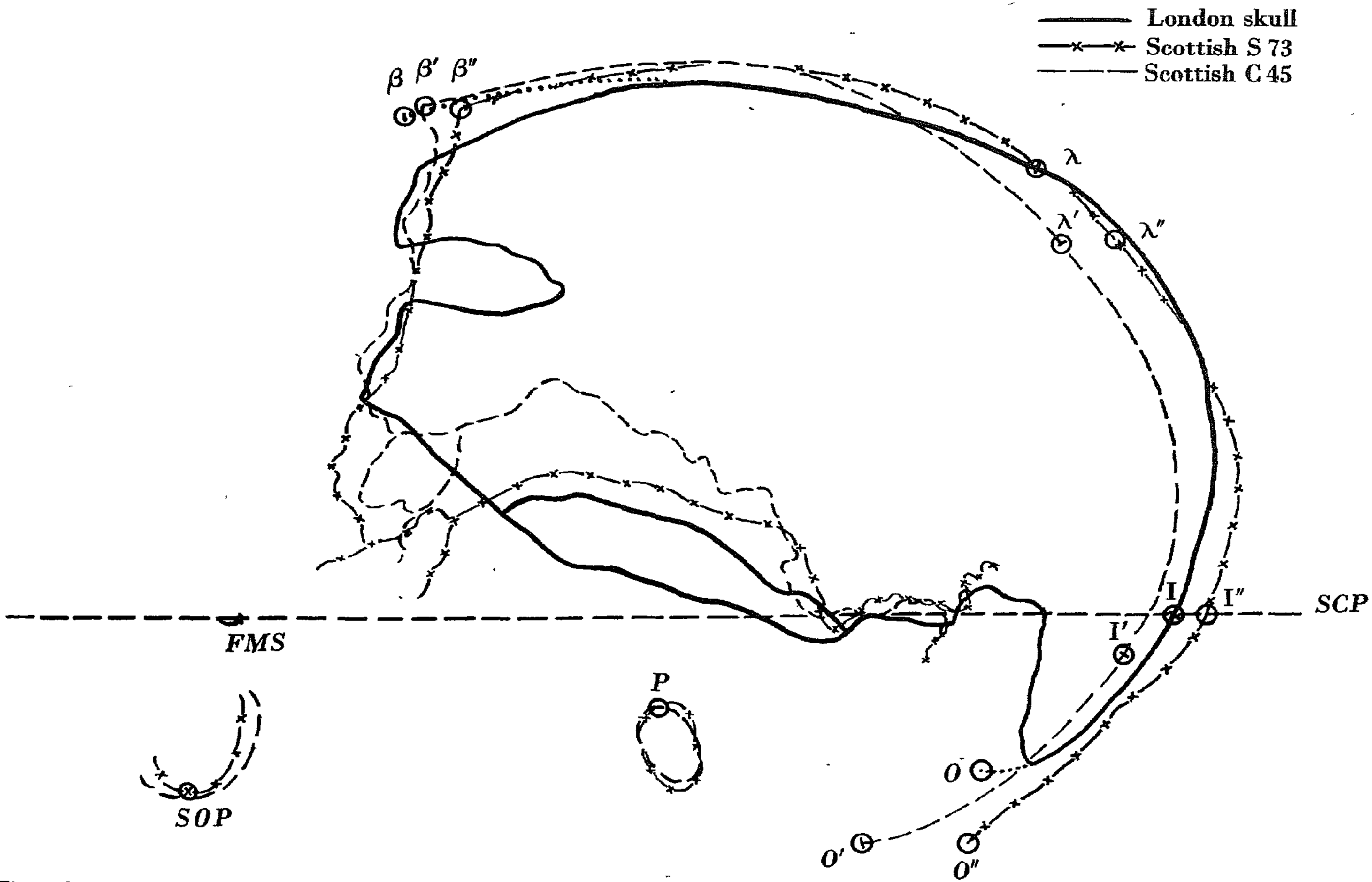


Fig. 4. Superimposition of the median sagittal tracing of the London skull on the corresponding tracings of two relatively low-vaulted members of the Scottish series, one (S 73) with an interparietal present, the other (C 45) with no interparietal. The skulls are oriented in the subcerebral plane, the asterions being made coincident.

supra-inial and sub-inial segments of the occipital section, were also measured directly with the steel-tape and callipers.

In addition to these new direct measurements and those which could be obtained from the incomplete orthographic projections, the measurements of various other cranial characters—including the maximum length, the maximum breadth, the biasterionic breadth and the cubic capacity—were available from previous records for each of the skulls comprising the two groups of the modern Scottish series. The values of the several cranial characters that it was possible to measure on the profile outline of the London skull fragment, and the dimensions of other characters which could be estimated with such a fair measure of probability that they may be supposed reasonable approximations to the true values, are shown in Table II. In the same table are given the mean values and standard deviations computed for a random group of forty-seven normal modern Scottish female skulls, and the same constants for ten specimens from the same series having interparietals or pre-interparietals. In the penultimate column all means of the Farringdon Street series* that are available for the characters used are given. This series has been chosen in preference to that from Whitechapel mainly because the mean values of certain characters relating to the occipital section are only available for the Farringdon Street skulls.

The measurements of most of the characters considered taken on a cast of the London skull and published by H. J. Friederichs† in his memoir on the specimen are also entered in the table, as in some cases they appear to differ greatly from those recorded by observers in this country who have studied the actual bones.

Before the measurements of the London skull fragment are compared in detail with the mean measurements of the two modern female series, attention may be drawn to differences in the two Scottish groups which appear to be associated with the presence of an interparietal. The figures in Table II show that the occipital arc is significantly longer and more curved and the occipital chord significantly longer, on the average, in the group with interparietals present than in the normal group. The average differences in length of arc and chord are 9 and 6 mm., respectively. The occipital curvature may be measured by the crude index, $100 \times$ occipital chord/occipital arc, or by Pearson's occipital index ($Oc.\ I = 100 \times \frac{S_3}{S_3'} \sqrt{\frac{S_3}{24\,(S_3 - S_3')}}$, where S_3 = occipital arc and S_3' = occipital chord). The latter index is preferable to the former, it measures the convexity of the occipital bone from the lambda to the opisthion, giving the ratio of the radius of curvature of the bone (supposing the curvature to be that of a circle, which is only roughly the case) to the occipital chord. As the radius of curvature shortens with greater convexity the index obviously decreases correspondingly.

* Taken from *Biometrika*, 1926, Vol. 18, p. 28.

† "Die morphologische Einreihung des 1925 in London City gefundenen paläolithischen Schädels." *Zeit. für Anat. und Entwick.* 1932, 98. Band, pp. 475–486.

TABLE II

Showing a Comparison of the Measurements of the London Skull (a) with those of a cast of it given by Friederichs (b), with a group of 47 Scottish female skulls without interparietals (c), with a group of 10 female skulls with interparietals belonging to the same series (e), and with a seventeenth-century London series (g)

	The London Skull		A random group of 47 normal Scottish female skulls		A group of 10 Scottish female skulls with interparietals		A series of seventeenth-century London skulls from Farringdon Street	$\frac{a-c}{d}$ ‡
	Young	Friederichs	Mean ± S.E.	σ	Mean ± S.E.	σ	Mean	
	a	*b*	*c*	*d*	*e*	*f*	*g*	*h*
Maximum length (L)	180*	—	178·4 ± 0·74	5·07	179 8 ± 0 99	3·12	181·6 (182)	—
Maximum breadth (B)	144?	152 (76 × 2)	135 5 ± 0 68	4 66	135·4 ± 1·77	5 58	135·7 (180)	1 8
Basio-bregmatic height (H')	120*	131 0	126·7 ± 0 64	4·40	128·7	—	122·5 (163)	—
Auricular height (OH)	—	114 0	107·3 ± 0 47	3·20	107·8	—	105 2 (69)	—
Height of cranial vault above subcerebral plane	90	—	98 2 ± 0 67	4 60	98·6 ± 1·10	3 47	—	−1·8
Height of squamous suture above subcerebral plane	22	—	30·5 ± 0 65	4 43	27·5 ± 1·44	4 55	—	−1·9
Biasterionic breadth (B''')	114	116 0	106 2 ± 0·60 (40)	3·81	107·0 (5)	—	—	2 1
Parietal arc (S_2)	105	130·0	122 7 ± 0 91	6 25	117·1 ± 1·83	5 80	123·8 (201)	−2·8
Occipital arc (S_3)	123	128·0	114·5 ± 0 96	6 59	123·5 ± 1·93	6·10	114·8 (182)	1·3
Lambda-inion arc ($S_{3.1}$)	82	81·5	60·9 ± 1·13	7·77	69 5 ± 2 06	6 50	[68·0]† (64)	2 7
Inion-opisthion arc ($S_{3.2}$)	41	46·5	53 7 ± 0 76	5·18	54 0 ± 1·81	5·74	[46 0] (64)	−2·5
Parietal chord (S'_2)	101	121·0	109·9 ± 0 69	4 71	106 3 ± 1·37	4 35	110·3 (201)	−1·9
Occipital chord (S'_3)	100	104·0	94·8 ± 0 70	4·78	100·6 ± 1·39	4·40	94·0 (184)	1 1
Lambda-inion chord ($S'_{3.1}$)	76	74·0	57 4 ± 0 97	6 62	64 6 ± 1 75	5 52	[63·0] (64)	2·8
Inion-opisthion chord ($S'_{3.2}$)	40	46·3	52 3 ± 0·73	5 04	52·4 ± 1 61	5·10	[45 0] (64)	−2 4
Height of parietal arc	12	21·5	22·7 ± 0 34	2·32	20 7 ± 0·84	2·67	[23 0]	−4·6
Height of lambda-inion arc	15	15 0	8 2 ± 0 32	2·21	10·2 ± 0 77	2·44	[10 0]	−3·1
Length of parieto-occipital segment of skull parallel to subcerebral plane	130	—	129·1 ± 0·76	5·19	130·3 ± 1·12	3·53	—	0·2
Lambda-inion-opisthion angle	117°	116°·0	121°·3 ± 0·56	3°·84	120° 1 ± 1·08	3 40	[120° 0]	−1 1
Lambda-inion subcerebral plane angle	77°	—	81°·9 ± 0·53 (30)	2° 91	81°·4 ± 1·14 (9)	3° 43	—	−1·7
Opisthion-inion subcerebral plane angle	40°	—	39°·9 ± 0 59 (30)	3° 25	39° 8 ± 1·01	3°·03	—	0 0
Length-breadth (cephalic) index (100 B/L)	80*	—	76·0	—	75 3	—	74·8 (167)	—
Length-height (altitudinal) index (100 H'/L)	—	—	71 0	—	71·6	—	67·9 (155)	—
Height-breadth index (100 B/H')	—	—	106·9	—	105·2	—	110·6 (158)	—
100 × cerebral height/basio-bregmatic height	—	—	77·5 ± 0·49	3 38	76·7	—	—	—
100 × cerebral height/auricular height	—	—	91 6 ± 0 54	3 70	91 7	—	—	—
Index of parietal curvature (100 S'_2/S_2)	96 2	93·8	89·8 ± 0 25	1·75	90 8 ± 0 40	1 26	89·1 (201)	3·7
Index of occipital curvature (100 S'_3/S_3)	81·3	81·2	82·9 ± 0 26	1 80	81·5 ± 0·58	1 84	81·9 (182)	−0 9

TABLE II (*continued*)

	The London Skull		A random group of 47 normal Scottish female skulls		A group of 10 Scottish female skulls with interparietals		A series of seventeenth-century London skulls from Farringdon Street	$\frac{a-c}{d}$ ‡
	Young	Friederichs	Mean ±S.E.	σ	Mean ±S.E.	σ	Mean	
	a	*b*	*c*	*d*	*e*	*f*	*g*	*h*
Pearson's occipital index	58·1	—	59·7 ±0·25	1·70	58·4 ±0·49	1·57	59·1 (181)	−1·0
Index of curvature of supra-inial arc ($100 \times S'_{3\cdot1}/S_{3\cdot1}$)	92·7	90·8	94·5 ±0·34	2·31	93·0 ±0·64	2·02	92·6 (64)	−0·8§
Index of curvature of infra-inial arc ($100 \times S'_{3\cdot2}/S_{3\cdot2}$)	97·6	99·6	97·5 ±0·25	1·69	97·1 ±0·53	1·66	97·8 (64)	0·1§
100 × parieto-occipital segmental length/maximum length (L)	—	—	72·4 ±0·33	2·29	72·5 ±0·73	2·30	—	—
100 × lambda-inion arc/occipital arc (S_3)	66·7	63·7	53·0 ±0·72	4·92	56·3 ±1·33	4·22	59·2	2·8
100 × inion-opisthion arc/lambda-inion arc	50·0	57·1	90·2 ±2·54	17·4	78·7 ±4·43	14·0	67·6	−2·3
Parietal arc height index	11·9	17·8	20·6 ±0·24	1·68	19·4 ±0·55	1·72	21·3	−5·2
Lambda-inion arc height index	19·7	14·4	14·1 ±0·38	2·61	15·6 ±0·82	2·59	15·9	2·2
100 × biasterionic B/parietal breadth (B)	79·2	—	78·3 ±0·44 (40)	2·84	78·3 (5)	—	—	0·3
100 × biasterionic B/parietal chord (S'_2)	112·9	—	96·5 ±0·80 (40)	5·05	101·2 (5)	—	—	3·3
100 × biasterionic B/occipital chord (S'_3)	114·0	111·5	111·6 ±1·01 (40)	6·41	105·1 (5)	—	—	0·4

* Measurements of the London Skull followed by an asterisk are merely estimated values.

† Measurements of the Farringdon Street skulls given in square brackets [ ] are estimated from the mean sagittal contour tracings published in *Biometrika*, 1926, Vol. 18

‡ It is important to emphasize that for all the characters, with two unimportant exceptions, in which the ratios in the last column of the table exceed 2, the mean values of the Scottish group with interparietals are less divergent from the corresponding measurements in the London skull than are the mean values of the Scottish group without this anomaly.

§ It is apparent in these cases that the distributions in the Scottish series are not approximately normal and hence that the ratios are criteria of a different kind from the others

Since it may be suggested that forty-seven skulls is a small number on which to base a standard of normality, it may be mentioned that the same striking differences are apparent on comparing the mean values of the characters under consideration in the interparietal group with the corresponding mean values derived from the whole series of female skulls, fully 370 in number. Closely analogous differences are shown, moreover, in comparing the mean values of the same three characters in the group of twelve male skulls showing interparietals or pre-interparietals with the corresponding mean values in the total series of fully 500 male skulls, though it should be stated that the difference in the mean occipital index in the male groups cannot be regarded as certainly significant as it is not quite twice its standard error.

On comparing the supra-inial and sub-inial segments of the occipital arc in the two female groups, it is seen that, as might be expected, the divergence which has been referred to is restricted to the upper region. The lambda-inion arc is

approximately 9 mm. longer and the lambda-inion chord approximately 7 mm. longer on the average in the group with interparietals than in the normal group. These excesses are statistically significant The difference in the index of curvature ($100 \times$ chord/arc) of the supra-inial segment in the two groups exceeds twice its standard error, indicating that the arc in the interparietal group may be regarded as sensibly more convex than in the normal group. In the sub-inial region, however, the respective measurements of the mean length and curvature of the arc and the mean length of the chord are practically identical in the two groups.

The parietal arc and chord and index of curvature ($100 \times$ chord/arc) also show differences between the two series. The absolute measurements are both sensibly greater and the ratio sensibly less on the average in the normal group than in the group with interparietals. As the upper occipital arc is, on the average, definitely longer in the anomalous than in the normal group, there would appear to be present a tendency to an inverse relationship in the lengths of the supra-occipital and parietal arcs. In the group of fifty-seven modern skulls, the correlation between the lengths of the parietal and lambda-inion arcs is $-0{\cdot}227 \pm 0{\cdot}126$.* Though the existence of a slight tendency to an inverse relationship in the extent of the arcs seems to be suggested by the coefficient, in view of the size of the standard error it cannot be said to be definitely significant for the measurements available. In the total group of 376 West Scottish female skulls there is, however, a small but statistically significant inverse association between the lengths of the total occipital and the parietal arcs, the coefficient being $-0{\cdot}182 \pm 0{\cdot}050$, a long occipital arc showing a slight tendency to be associated with a short parietal arc and vice versa. In the long Egyptian series of female skulls of the 26th to 30th dynasties, the compensatory relationship in the two corresponding cranial arcs is more emphasised than in the Scottish series, the coefficient of correlation between the lengths of these two segments of the sagittal arc being $-0{\cdot}342 \pm 0{\cdot}037$.†

The definite differences observed in the mean measurements for the two groups of skulls from the same relatively homogeneous Scottish series—that with and that without interparietals—suggest strongly that the relatively long occipital bone and the relatively short parietal bone, which are features of the London skull fragment that have been noted by various observers, are really largely dependent on the fact that it has a pre-interparietal bone.

We pass now to a comparison of the measurements of the characters in the London skull fragment with the corresponding mean measurements in the modern Scottish female series and in the seventeenth-century series of female Londoners. A brief scrutiny of the figures in Table II shows the close correspondence that obtains in general between the corresponding average values in these two series. The comparison will be mainly concerned with the mean values

* The symbol $\pm$ denotes standard errors throughout this paper.

† *Biometrika*, 1924, Vol. 16, p. 361.

in the Scottish series, and only when these differ appreciably from the values in the London series will particular reference be made to the latter. The deviation of the measurement of any character in the London skull from the mean value of the corresponding character in the normal Scottish series was divided by the standard deviation of the character for this series. The ratios thus obtained vary considerably in value. When the ratio is 1·0, then 1 skull in 6 has the characteristic more emphasized in the given direction than it is in the London skull; if 1·5 then 1 skull in 15; if 2·0 then only 1 in 45, if 2·3 then only 1 in 93; if 2·5 only 1 skull in 160. In statistical judgments, a characteristic is not usually regarded as individually remarkable unless it exhibits a difference *in the given direction* that is not equalled or exceeded more than once in about 45 times or more, that is, when the ratio equals at least 2.

Applying this test to the various characters of the London skull and beginning with the occipital region, we find, as shown in the last column of Table II, that though the measurements of the occipital (lambda-opisthion) arc, the occipital chord and the occipital indices of curvature—both the crude one ($100 \times$ chord/arc) and Pearson's—in the specimen deviate to some extent from the mean values in the normal Scottish female series, yet the divergences are not so great that the characters can be considered exceptional for this series. The supra-inial arc curvature and infra-inial arc curvature in the cranial fragment are also of such a degree as might be likely to occur not infrequently in the normal Scottish female, but the supra-inial arc and chord are more extensive, and the infra-inial arc and chord less extensive than might reasonably be expected to appear except rarely in this modern type. Agreement with the Scottish skull type in these respects is definitely closer, however, when interparietals are present, as is shown by the fact that the lengths of the occipital arc and the occipital chord, and the curvatures of the total occipital and supra-inial arcs in the London specimen, practically coincide with the corresponding mean values in the anomalous or interparietal Scottish group. Dimensions of the supra-inial arc and chord corresponding in extent to those found in the London skull might also be expected to occur with greater frequency in the Scottish group with interparietals than in the normal group.

In view of the relatively small extent of the sub-inial region of the London skull in the sagittal plane, it should be mentioned that some allowance has been made in the estimated measurements of the occipital bone for the small area that has been detached and lost at the opisthion. The disproportion in length of the supra-inial and sub-inial segments of the occipital arc in the London skull is very striking. In this specimen, as will be seen from reference to Table II, the ratio of the inion-opisthion arc to the lambda-inion arc is 50 per cent., as compared with corresponding values of approximately 90 and 70 per cent. in the Scottish and Farringdon Street series, respectively. A possible explanation of this feature may be provided from analogy with a relationship which is found in the Scottish

skull. In the Scottish series of fifty-seven skulls there is a very definite inverse association between the lengths of the supra-inial and sub-inial arcs, as is shown by a coefficient of correlation of $-0{\cdot}454 \pm 0{\cdot}105$, a long supra-inial segment tending to be associated with a relatively short infra-inial segment.

The London skull is also incomplete in the region of the bregma, but the approximate position of this anthropometric point or landmark can be indicated by continuing the outline tracing forward till it meets the line of the coronal suture. The parietal arc measured from the lambda to the bregma determined in this way is characteristically less than the corresponding measurement in the normal Scottish series (which is identical with that in the recent Londoners). The parietal curvature, as measured by the index, $100 \times$ parietal chord/parietal arc, is also characteristically less, and the parietal chord in the London skull is of a length which might reasonably be expected to occur occasionally in the modern female. From analogy with the definite tendency to an inverse relationship in occipital and parietal arc lengths, which has been shown to exist in both the Scottish and Egyptian female skulls, the occurrence of a relatively short parietal arc in the London skull might almost be expected, as its occipital arc is relatively long and the supra-inial segment of that arc exceptionally long.

It should be mentioned here that, while the measurements cited for the lengths of the sagittal parietal arc and chord are almost identical with those estimated from the profile drawing of the London skull given by Sir Arthur Keith,* they differ greatly from those for the corresponding characters given by Friederichs in his memoir on the specimen. This observer estimates the parietal arc length to be 130 mm. and the chord 121 mm. Reasoned consideration of the probable form and dimensions of the skull based on the fragment that is preserved has convinced the writer that measurements of this order for the parietal bone are obviously beyond the range of what is not merely probable but possible. The same criticism seems to be applicable to some of the estimates of measurements of other principal characters of the London skull recorded by Friederichs and cited in Table II. Half the maximum breadth is tabulated by him as 76 mm., which multiplied by 2 would give a maximum width of 152 mm., though in a later table the width is given as 144–146 mm. His estimates of the basio-bregmatic height and auricular height as 131 and 114 mm., respectively, appear to be much in excess of the most probable values.

Certain other characters in the London skull may now be compared with the corresponding characters in the modern series. A notable feature of the cranial fragment on which much emphasis has been laid is the apparent lowness of the vault. The highest point of the vault—i.e. the vertex—lies according to Sir Arthur Keith only 90 mm., approximately, above the level of the subcerebral or basal plane. The average height of the summit of the vault above the same base line in the Scottish random series of forty-seven female skulls is 98·2 mm. As

* *New Discoveries relating to the Antiquity of Man*, 1931, p. 446 *et seq.*

the same height of the London skull does not deviate from this average value by twice the standard deviation of the elevation in the Scottish series it cannot be considered a value so low that it is unlikely to appear occasionally in this series; indeed, in the group of fifty-seven skulls (forty-seven normal and ten anomalous) three specimens are found in which the subcerebral heights of the vault are 88, 89 and 90 mm., respectively, while seven other members of the series do not exceed 93 mm. in altitude. The height of the cranial vault above the subcerebral plane in the London skull cannot, therefore, be described as extremely low in comparison with that found in the modern type.

In the sample of Scottish female skulls, the highest point on the squamous suture lies on the average 30·5 mm. above the level of the subcerebral plane. In the London fragment the corresponding height is 22 mm. As the standard deviation in the Scottish series is 4·4 mm. the height of the suture in the London fragment cannot be considered a value which would be exceptional for the Scottish series; indeed, the actual elevation of the suture above the basal plane in this series varies from 18 to 40 mm. Sir Arthur Keith notes as a point of contrast between the London skull and a modern "river-bed" type of skull with which he compared it, the greater rapidity with which in the former the lower or squamous border of the parietal rises as it passes forward. A brief survey of the outline tracings of the modern Scottish series supplies ample evidence that this feature is extremely variable even in such a homogeneous group, and that a wide range in degree of inclination is found; in some cases the direction of the suture in its posterior part almost approximates to the vertical.

From a comparison of a low-vaulted skull from the "river-bed" series with the London skull by superimposition of the profile outlines, Sir Arthur Keith came to the conclusion that one of the features in which the London skull resembles the skull from Piltdown, but differs from specimens of the modern type, was that in the first-named specimen the sub-inial or nuchal part of the skull descended in a more vertical direction. In reaching this conclusion, he does not appear, however, to have made adequate allowance for the normal range of variation in this feature in the modern skull. A rough estimate of the degree of flexion of the lower sub-inial segment of the occipital bone on the supra-inial segment may be obtained by measuring the lambda-inion-opisthion angle, i.e. the angle enclosed by the chords of the two segments of the occipital bone. The size of this angle is not a very reliable index of the degree of flexion, however, as it is influenced to such an extent by the variable position of the inion. In the Scottish series it ranges in value from 111 to 127° with an average of 121°. In the London skull the angle is 117°, which is well within the limiting values of the modern group.

The acuteness of the forward flexion of the sub-inial segment may also be estimated in some measure by the size of the angle between the inion-opisthion chord and the line denoting the subcerebral plane. This angle in the London skull

is 40°. In this skull, the internal occipital protuberance, unlike the arrangement in the Neanderthal type, coincides in level with the inion. In the thirty specimens from the Scottish female series in which the subcerebral plane passes approximately through the inion the average value of the angle is 39°·9 and the range of variation from 33 to 45°. These comparisons seem to indicate that in regard to the curvature of the lower occipital region, the London skull cannot be held to differ in a significant degree from the specimens of modern type.

It has already been mentioned that one of the main differences between the fragment of the London skull which has been preserved and the corresponding part of the Neanderthal type of skull is seen in the lower part of the occipital region. In the latter type the sub-inial part of the occipital bone does not continue downward the line of curvature of the upper segment but is bent somewhat abruptly forwards at the inion corresponding to the flattened form of the cerebellum.

On account of the defect in the right parietal region of the London skull it is not easy to determine with precision the maximum parietal breadth, but it is probably at least 144 and possibly 146 mm. Sir Arthur Keith gives the measurement as 140 mm., but this would appear to be an under-estimate, as he states that the width of the endocranial cast is 136 mm. and that the thickness of the skull wall varies from 5 to 7 mm. The thickness of the cranial wall is at least 5 mm. at the widest part of the parietal region. In the Scottish series of forty-seven female skulls the average maximum breadth is 135·5 mm. The standard deviation of the breadth in the group is 4·7 mm., so that a cranial width of 144 mm. is a measurement that might reasonably be expected to occur occasionally in such a modern series; indeed, in the series are present two specimens with maximum parietal breadths of 145 and 146 mm., respectively. It is important to note that the maximum parietal breadth of the London skull is found at a point well forward on the parietal bone, and not relatively far back as occurs in skulls of Neanderthaloid type.

A defect in the left asterionic region makes it difficult to estimate with accuracy the biasterionic diameter of the London skull, but it appears to be approximately 114 mm. In forty skulls of the modern Scottish female series the mean biasterionic breadth is only 106·2 mm. An individual measurement as great as 114 mm. in this series must be regarded as rather exceptional. The proportions which the biasterionic diameter in the London skull bears to the maximum parietal breadth (B) and to the length of the occipital chord (S_3'), respectively, are, however, in fairly close agreement with the corresponding average indices in the modern Scottish female series.

Having considered the characters of the London skull fragment that can be measured with a reasonable approach to accuracy, it may be of interest to place on record some observations on the probable length, form and cubic capacity of the cranium when complete. An estimate of the original maximum length may

be obtained from the different dimensions of the fragment that are available. Prof. Karl Pearson* found in his very long series of Egyptian female skulls that of the three segments of the sagittal arc—frontal, parietal and occipital—the last (S_3) shows the highest correlation with the maximum length (L) and would give the most accurate prediction of this character. In the Scottish female series of 375 skulls, the correlation between the maximum length (L) and the length of the occipital arc (S_3) is $0{\cdot}501 \pm 0{\cdot}039$, which is of much the same order as the value ($r = 0{\cdot}442 \pm 0{\cdot}034$) found in the long Egyptian female series. The linear regression equation expressing maximum length in terms of occipital arc length in the Scottish series is

$$L = 0{\cdot}38S_3 + 136{\cdot}00,$$

with a standard error of prediction of 4·5 mm. If this equation be assumed to be applicable to the London skull, in which the length of the occipital arc equals 123 mm., the maximum length may be considered to be approximately 183 mm.

The length of the parieto-occipital segment of the London skull from the occipital contour line to the approximate position of the coronal suture, near the point where this suture is crossed by the lower temporal line, measured in a direction parallel to the subcerebral plane, is 130 mm. The correlation coefficient between the length of this parieto-occipital segment ($P.O.L.$) and the glabella-occipital length (L) in the Scottish series of fifty-seven female skulls is $0{\cdot}543 \pm 0{\cdot}093$. The regression equation expressing maximum length in terms of parieto-occipital length is

$$L = 0{\cdot}52P.O.L. + 111{\cdot}14,$$

with a standard error of prediction of 4 mm. Assuming this equation to be applicable to the London skull, in which the parieto-occipital segment is 130 mm., the predicted maximum length would be 179·3 mm. or approximately 180 mm. This estimate of length is probably nearer the true value than that based on the regression formula for the length of the occipital arc. The maximum cranial length may be estimated in a simpler way. In the Scottish female series the average length of the parieto-occipital segment is 129 mm., i.e. almost identical with the corresponding measurement in the London skull. The length of this segment expressed as a proportion of the maximum length has in the series an average value of 72·4 per cent. with a range of variation from 67 to 77 per cent. If we assume that the parieto-occipital segment in the London skull is 72·4 per cent. of its maximum length, the maximum length of the skull would again be 179·6 mm. On the assumption that the parieto-occipital length just described is on the average approximately 70 per cent. of the total length in modern skulls, Sir Arthur Keith suggests that the probable length of the London skull was 185 mm. On the assumption that the mean ratio in the Scottish modern female series may be considered to represent approximately the relationship in the

* *Biometrika*, 1924, Vol. 16, Table V, facing p. 348.

London skull, the maximum length in the latter may be held to be most probably about or just under 180 mm., but it should be borne in mind that the actual length may be appreciably greater or less than this estimate. Let us consider for a moment the limiting values found for the ratio $100 \times$ parieto-occipital length/greatest length, in the homogeneous Scottish group. A value of 67 per cent. for the ratio would correspond to a total length of 194 mm., while a ratio of 77 per cent. would postulate a length of 169 mm. Though, as is known from the general form and extent of the fragment preserved, neither of these limiting lengths is probable, indeed possible, for the London skull, yet they serve to indicate that the most probable length of 180 mm. may be appreciably in excess or defect of the true value.

A maximum length of 185 mm. and a maximum breadth of 140 mm., as estimated by Sir Arthur Keith, give a cephalic index of 75·7, which practically coincides with the mean value for the complete Scottish female series. A maximum length of 180 mm. and a maximum breadth of 144 or 146 mm. give a cephalic index of 80 or 81, but even such a relatively high index cannot be considered one that is unlikely to be found occasionally in the Scottish female skull.

Having considered the probable maximum length of the skull, we may refer briefly to its probable cubic capacity. Various formulae have been computed—notably by Pearson and Lee* and Hooke†—for the purpose of determining the cranial capacity from the absolute linear measurements of the skull length, breadth and height (auricular or basio-bregmatic) or their product ($L \times B \times H'$ or $L \times B \times OH$). As it is possible that in other fragmentary skulls, like that found in London, the height of the vault from the subcerebral plane may be determinable when neither basio-bregmatic nor auricular height can be ascertained, the linear regression equation expressing the cranial capacity in terms of the product of the absolute measurements of length, breadth and subcerebral height has been calculated, based on the data which are available for the fifty-seven Scottish female skulls. In this series the correlation between the subcerebral height (H'') and the cubic capacity (C) is almost as high as that between the maximum length and the cubic capacity, the respective coefficients being $0·504 \pm 0·099$ and $0·529 \pm 0·095$. The equation is as follows:

$$C \text{ (in c.c.)} = 0·000465 \times (L \times B \times H'') + 228·84.$$

The standard error of prediction is 60 c.c. This formula may be applied to estimate the cubic capacity of the London skull. Using the values of the three linear dimensions which have been suggested as most probable as a result of the present study—viz. length $(L) = 180$ mm., breadth $(B) = 144$ mm. and subcerebral height $(H'') = 90$ mm.—the estimated capacity is 1314 c.c. From the

* *Phil. Trans.* 1899, Vol. 196 A, pp. 225–264.

† *Biometrika*, 1926, Vol. 18, pp. 33 and 34.

linear dimensions given by Sir Arthur Keith—viz. length = 185 mm., breadth = 140 mm. and subcerebral height = 90 mm.—the predicted cubic capacity is exactly the same. This estimate is rather greater than the cubic content, viz. 1260 c.c., cited by this author as determined by the application of the formula of Pearson and Lee, but the difference does not exceed the standard error of prediction in using the equation which is given. If we may consider the cubic capacity as being probably in the region of 1300–1320 c.c. we have an estimate which does not diverge appreciably from the mean cubic capacity for 376 female skulls in the Scottish series, viz. $1329 \pm 3{\cdot}6$ c.c.

Though it must be admitted that the prediction of the maximum length of the skull from the antero-posterior extent of its parieto-occipital segment can only be considered a rough approximation to the true value, the cubic content computed from the product of the length thus estimated and the other two linear dimensions that can be estimated with a greater degree of accuracy suggests that it does not diverge greatly from the average cranial capacity found in a group of modern female skulls of a type closely related to that found in seventeenth-century Londoners.

A survey of the mean values of the several cranial characters in Table II shows that one feature in which the seventeenth-century London female differs from the modern Scottish female and approaches more nearly to the condition found in the London skull fragment is the relative proportion of the two segments of the occipital bone The lengths of the complete occipital arc and chord and the degree of curvature of the occipital arc as a whole in the recent Londoner are almost identical with the corresponding characters in the normal Scottish series. The lengths of the supra-inial segment of the occipital arc and the supra-inial chord and the degree of curvature of the supra-inial segment in the seventeenth-century Londoner are, on the other hand, in close agreement with the corresponding characters in the Scottish group with interparietals, and like these sensibly divergent from the corresponding values in the normal group, whereas the infra-inial arc and chord in the seventeenth-century Londoner are definitely shorter than the corresponding characters in either of the Scottish groups and lie nearer to the estimated values of the London skull than to these means.* The difference in the relative proportions of the two segments of the occipital arc in the two groups is shown clearly by the tabulated values of either of the two ratios: $100 \times$ inion-opisthion arc/lambda-inion arc and $100 \times$ lambda-inion arc/occipital arc.

Some of the main points brought to light from a study of the figures given in Table II to which reference has been made in the preceding section of the text may be summarized briefly here. The standard deviations of the several charac-

* It is possible that the disagreement between the Scottish and London series in the measurements taken from the inion may be due partly to differences in the way in which this point was located by the different observers.

ters in the Scottish series may be accepted as rough measures of the variabilities that might be expected in the corresponding characters in the London skull were it a member of a series, and give some indication of the error that may be made in assuming the specimen to be an average representative of its type. From the differences that are observed in the mean measurements of the two Scottish female groups—the normal and that with interparietals—there is reasonable ground for the inference that the presence of this anomalous bone in the London skull has probably had a definite influence in determining the rather unusual extent and curvature of its occipital region. Of the differences that are observed in the measurements of the characters in the London skull and the corresponding mean values of the Scottish series, only a few are of such a magnitude as to indicate that the measurements in the specimen may reasonably be considered exceptional for the Scottish sample. Amongst these, the principal are the relatively long supra-inial and short infra-inial segments of the occipital arc and the relatively short and flat parietal arc. It is not improbable, however, that in the London skull, the short parietal arc, as well as the relatively short infra-inial arc, may be to some extent compensatory to the long supra-inial arc which is associated with the presence of the pre-interparietal.

In height of cranial vault above the subcerebral plane, the London skull cannot be said to be exceptionally low for a modern skull of the Scottish type, and in fullness of curvature of the lower occipital region it appears to be not greatly divergent from, if not in close agreement with, the latter.

Comment is made on the estimates of the dimensions of some of the characters cited from Friederichs' memoir on the London skull. His measurements were taken on a cast. Those given for the lengths of the parietal arc and chord and the basio-bregmatic height, at least, appear from the form and dimensions of the actual fragment of skull that is preserved to be very improbable, if not impossible, approximations to the true values.

(*d*) *Comparison with the Neanderthal Female Skulls from Gibraltar, La Quina and Saccopastore* (*Rome*). In view of the alleged Neanderthaloid features in the London skull fragment, to which reference has already been made, and its alleged affinities with the Piltdown skull, which have been discussed by Sir Arthur Keith,* it seemed to be of interest to compare in detail with the measurable characters of the London skull the measurements of the corresponding characters in the three skulls of Neanderthal type which are generally admitted to be of the female sex, namely the Gibraltar I, the La Quina (adult) and the Saccopastore, and also those in the Piltdown skull. A brief account will also be given of the corresponding characters in the recently discovered Swanscombe skull.

Unfortunately, on account of deficiencies in the respective crania, some of the characters that have been under review cannot be measured in such a manner as to warrant any considerable degree of confidence in the probable accuracy of the

* *New Discoveries relating to the Antiquity of Man*, 1931, p. 446.

estimates obtained. Such measurements of characters have been tabulated, however, as may probably be deemed at least fair approximations to the true values, and a few comments will be made on their relationships to the corresponding measurements in the London skull and, incidentally, to those in the modern Scottish series; these measurements are given in Table III.

The measurements of the La Quina and Gibraltar specimens which are given in Table III have been taken from the measurements and contour tracings published by Dr G. M. Morant.* The absolute measurements of the arcs and chords of the Saccopastore skull were very kindly supplied by Prof. Sergio Sergi at the request of the late Sir Grafton Elliot Smith. We are greatly indebted to Prof. Sergi for permission to use these measurements before his final memoir on this very important skull is published. The measurements of the Piltdown skull are taken from the reconstruction made by Prof. Elliot Smith with the assistance of Dr John Beattie. The measurements of the Swanscombe skull were taken from the actual specimen by the courtesy of Mr A. T. Marston.

In preliminary accounts of the Saccopastore skull Sergio Sergi† states that among the other Neanderthal skulls, the specimen from Gibraltar (Gibraltar I) is the one that approaches nearest the Saccopastore in general dimensions and in morphological type. As the Saccopastore skull is not only the specimen most nearly complete but also the one that in its general dimensions is nearest the London skull fragment, the main comparison will be made between the characters of the London skull and this specimen as representing the Neanderthal female type.

In maximum length (L), maximum parietal breadth (B), biasterionic breadth (B''') and cephalic index ($100\,B/L$) the Saccopastore skull is not very different from the London skull fragment. With a basio-bregmatic height of 109 mm., however, the vault of the former is apparently much lower than what may be considered the most probable height for the London skull (*ca.* 120). The relative shortness of the parietal segment (S_2) and relatively great extent of the occipital segment (S_3) of the sagittal arc in the London skull have already been commented upon, and their probable association with the presence of a pre-interparietal bone in this fragment discussed. Although the parieto-occipital segment of the sagittal arc (S_2+S_3) in the London skull is exactly equal to the corresponding segment in the Saccopastore skull, both being 228 mm., there exists a definite disproportion in the length of this segment formed by the parietal and occipital bones in the two specimens. The London skull parietal arc, though relatively short, is 105 mm., but that in the Saccopastore is 86 mm., i.e. 19 mm. less, whereas the occipital arc of the former is only 123 mm. as compared with 142 mm. in the Saccopastore specimen.

* "Studies of Palaeolithic Man. II." *Annals of Eugenics*, 1927, Vol. 2.

† (i) "Le Crâne Néanderthalien de Saccopastore (Rome)." *L'Anthropologie*, 1931, Vol. 41, p. 241. (ii) "Some Comparisons between the Gibraltar and Saccopastore Skulls." *Proc. 1st Internat. Congress Preh. and Protoh. Sciences, London*, 1932, pp. 50–52.

TABLE III

Showing the Measurements of the Available Characters in the London Skull in comparison with the Corresponding Measurements in (1) the three Neanderthal female Skulls: Gibraltar, La Quina and Saccopastore, (2) the Piltdown Skull and (3) the Swanscombe Skull

Characters	Symbols	The London skull	The Gibraltar I skull*	The La Quina skull (adult)*	The Saccopastore skull†	The Piltdown skull‡	The Swanscombe skull§
Maximum length	(L)	180‖	192·5	204·2	181 ?	194‖	180‖
Maximum parietal breadth	(B)	144	149·0	138·3	142·	150	142
Basio-bregmatic height	(H')	—	120·2 (Sergi)	—	109	—	126
Biasterionic breadth	(B''')	114	118·0 ?	112·4 ?	115	123 ?	121
Parietal arc	(S_2)	105	—	107·0	86	120	114
Occipital arc	(S_3)	123	110·0	—	142	125	115
Lambda-inion arc	$(S_{3\cdot1})$	82	65·0	66·3	90	70¶	63
Inion-opisthion arc	$(S_{3\cdot2})$	41	46·0	—	52	55	52
Parietal chord	(S'_2)	101	—	103·0	83	110	106
Occipital chord	(S'_3)	100	85·0	—	103	102	94
Lambda-inion chord	$(S'_{3\cdot1})$	76	60·0	57·8	80	65	58
Inion-opisthion chord	$(S'_{3\cdot2})$	40	44·6	—	50	53	51
Cephalic index	$(100\,B/L)$	80·0‖	77·5	67·7	78·5	78·0‖	79·0‖
Altitudinal index	$(100\,H'/L)$	—	—	—	60·2	—	70·0‖
Index of parietal curvature	$(100\,S'_2/S_2)$	96·2	—	96·3	96·5	91·7	93·0
Index of occipital curvature	$(100\,S'_3/S_3)$	81·3	77·3	—	72·5	81·6	81·7
Pearson's occipital index	$(Oc.\ I.)$	58·1	55·4	—	53·7	58·3	58·4
Index of curvature of supra-inial arc	$(100 \times S'_{3\cdot1}/S_{3\cdot1})$	92·7	92·3	87·2	88·9	92·9	92·1
Index of curvature of infra-inial arc	$(100 \times S'_{3\cdot2}/S_{3\cdot2})$	97·6	97·0	—	96·2	96·4	98·1
Lambda-inion arc/occipital arc	$(100\,S_{3\cdot1}/S_3)$	66·7	59·1	—	63·4	56·0	54·8
100 Biasterionic B/occipital chord	$(100\,B'''/S'_3)$	114·0	138·8	—	111·7	120·6	128·6
100 Biasterionic B/maximum parietal breadth	$(100\,B'''/B)$	79·2	79·2	81·3 ?	81·0	82·0	85·2
100 Biasterionic B/parietal chord	$(100\,B'''/S'_2)$	112·9	—	109·2 ?	138·6	111·8	114·2
Occipital (lambda-inion-opisthion) angle	—	117°·0	107°·0	—	—	125°·0	114°·0

* Measurements of the Gibraltar and La Quina skulls are taken from a memoir by G. M. Morant, *Annals of Eugenics*, 1927, Vol. 2, pp. 318–381.

† Measurements supplied by Prof. Sergio Sergi.

‡ From the reconstruction by Sir Grafton Elliot Smith and Dr John Beattie.

§ Measurements taken by permission of Mr A. T. Marston.

‖ Measurements indicated thus (‖) are merely estimated values.

¶ Estimated to be 75 mm. by Sir Arthur Keith.

The phenomenally short parietal arc in the Saccopastore skull is undoubtedly associated with the presence in the region of the lambda of several accessory ossicles which, if taken together, would correspond in position and extent with a fair-sized pre-interparietal bone. The point which Sergi has identified as the 'real lambda' and to which he has measured the arcs is L_1, the point at which the interparietal (or sagittal) suture meets the lambdoid ossicle most anteriorly placed. This is also the point in the middle line, or in the sagittal plane, at which the lines that correspond to the directions of the lateral parts of the lambdoid suture meet when prolonged medially. When Wormian bones are found in the region of the upper part of the occipital squama this, as described in Martin's *Lehrbuch*, is the conventional method of defining the lambda for the purposes of measurement. Sergi states that there are at least four other possible lambdas placed more posteriorly in, or near, the middle line between the ossicles, L_2, L_3, L_4 and L_5. The length of the arc between L_1 and L_5 is stated to be 35 mm.; Sergi has discussed the subject in his paper entitled: "Ossicini fontanellari della regione del lambda nel cranio di Saccopastore e nei crani neandertaliani."* He also gives in this paper a natural-size drawing of the occipital view of the skull, showing the accessory bones. The measurement given for the occipital arc in the Saccopastore skull is 142 mm. and that for its supra-inial segment 90 mm. That these may be considered exceptionally large measurements is clearly illustrated by comparing them with the measurements of the corresponding characters in the La Chapelle-aux-Saints skull—the largest male skull of the Neanderthal type yet described. In this specimen the occipital arc is about 117 mm and the supra-inial arc 74 mm. In an exceptionally large modern skull with a capacity of 2450 c.c. and a pre-interparietal bone present the length of the occipital arc was found to be 153 mm., merely 11 mm. more, and the length of the supra-inial arc 86 mm., i.e. 4 mm. less than the value for the specimen from Rome.

Though the parietal arc and chord in the Saccopastore specimen are unusually short, the index of curvature of the parietal arc is much the same as that estimated for the London skull and as that recorded for the La Quina skull.

The curvature of the occipital bone as a whole in the Saccopastore skull, as shown by the values of Pearson's occipital index and the chord-arc index, $100 \times S_3'/S_3$, is much greater than in the London skull and also rather greater than in the Gibraltar skull. The curvature of the supra-inial segment of the occipital arc in the Saccopastore, as in the La Quina specimen, is greater than in the London, though the Gibraltar agrees closely with the last-named specimen in this feature. The curvature of the infra-inial arc in the Saccopastore does not differ greatly from that in the London skull which also agrees closely with that in the Gibraltar skull.

As already mentioned, the biasterionic breadth in the Saccopastore specimen

* *Rivista di Antropologia*, 1934, Vol. 30.

is almost identical with the corresponding measurement in the London skull. In two of the three indices tabulated of which the biasterionic breadth is a component the Saccopastore skull does not differ greatly from the London skull. Thus 100 B'''/S_3' in the Saccopastore is 112, in the London skull 114; 100 B'''/B in the Saccopastore is 81, in the London skull 79. The third index 100 B'''/S_2' in the Saccopastore, on the other hand, is 139, whereas the corresponding index in the London skull is 113. This divergence is wholly due to the very short parietal chord (S_2') in the Saccopastore specimen.

The comparison of the London skull with the Gibraltar and La Quina specimens will be dealt with more briefly. The Gibraltar skull is much longer and rather broader, the La Quina skull much longer and narrower than the London. In biasterionic breadth, as in maximum parietal breadth, the London skull lies intermediately to these two Neanderthal skulls.

The lengths of the parietal arc and chord are not available for the Gibraltar skull, but in the La Quina specimen they are almost the same as the corresponding characters in the London skull, and the indices of curvature of the parietal bone (100 S_2'/S_2) in the two are practically identical. In the pronounced flatness of the parietal segment of its sagittal arc the London skull undoubtedly presents a definite Neanderthaloid feature.

In the Gibraltar skull the length of the occipital arc (S_3) is much less than the corresponding arc in the London skull, and still less than that in the Saccopastore. The curvature of the occipital arc in the Gibraltar is rather greater than in the London. The lambda-inion arc and chord in the Gibraltar specimen are also much less than those shown in the London skull, but they are reduced in proportion as the curvature of the arc agrees closely with that in the London skull. In the La Quina specimen the lengths of the lambda-inion arc and chord do not differ notably from the corresponding measurements in the Gibraltar skull but the curvature of the arc is rather greater The inion-opisthion arc cannot be measured in the La Quina specimen, but in the Gibraltar skull the arc and chord are about 5 mm. greater than the corresponding characters in the London skull; the ratio of chord to arc, i.e. the index of curvature, in the two skulls is thus almost identical.

Turning to the three indices of which the biasterionic breadth is a component, two only are calculable for each of the skulls from Gibraltar and La Quina. The biasterionic breadth as a percentage of the maximum parietal breadth in the London skull is identical with the corresponding ratio in the Gibraltar skull, and is but two units less than that in the La Quina specimen. The biasterionic breadth-parietal chord index (100 B'''/S_2') is not available for the Gibraltar specimen but is measurable in the La Quina skull. In this skull it is rather less than in the London specimen, but the difference is less than four units. The biasterionic breadth-occipital chord index (100 B'''/S_3') is not available for the La Quina specimen. In the Gibraltar skull it is much greater than in the London skull.

The difference is mainly due to the great difference in the length of the occipital chord (S_3'); this is 15 mm. shorter in the Gibraltar than in the London skull.

The greater curvature of the occipital bone in the Gibraltar skull than in the London is also brought out by the size of the occipital (lambda-inion-opisthion) angle. In the London skull the angle is 117° but in the Gibraltar skull it is 10° less.

(e) *Comparison with the Piltdown Skull.* It must be emphasized that no claim is made that any measurements made on the reconstructed Piltdown skull are to be considered more than rough approximations to the true values of the characters. If resemblances between its general form and that of the London skull can be recognized by superimposition of their outline tracings in different planes, it might reasonably be expected that some indication of these resemblances will be indicated by a comparison of the measurable characters and their relative proportions. As will be seen by reference to Table III, both the extent and curvature of the occipital segment of the sagittal arc are in close agreement with the corresponding characters in the London skull, but not closer than those for the mean skull in the Scottish anomalous group (Table II). The supra-inial segment of the occipital arc in the Piltdown skull, though it is definitely less extensive than that of the London skull, shows a similar degree of curvature. The infra-inial arc, on the other hand, appears to be definitely longer than, though also little different in curvature from, that in the London skull. The ratio of the supra-inial segment of the occipital arc to the arc as a whole (56 per cent) in the Piltdown thus differs considerably from that in the London skull; it is in fair agreement, however, with the average proportion found in the normal modern Scottish series, practically coinciding with that in the anomalous Scottish group and differing by not more than three units from the corresponding character in the Gibraltar skull.

The angle between the inion-opisthion chord and the line of the subcerebral plane, the size of which has been taken as a crude index of the fullness of curvature in the lower occipital region, is about 42°, i.e. the same as in the London skull. It must be recollected, however, that in the latter specimen the inion-opisthion chord is relatively short. In the Scottish skull, for which the mean value of the inion-opisthion chord agrees closely with that in the Piltdown skull, the corresponding angle is also practically the same as in the London.

The parietal segment of the sagittal arc in the Piltdown skull in extent and curvature seems to correspond closely with the corresponding mean values in the modern Scottish series; it is not relatively short and flat as in the London skull. The biasterionic breadth in the Piltdown skull appears to be definitely in excess of the corresponding measurement in the London skull, and the proportions borne by this diameter to the maximum parietal breadth and to the occipital chord, respectively, in it are also appreciably greater than the corresponding ratios in the London. The ratio of biasterionic breadth to parietal chord is, however, almost the same in the two specimens.

So far as available measurements of characters and their proportions give any material indication of relationship, the London skull does not appear to show any closer affinity with the Piltdown type than with the modern Scottish type.

(*f*) *Comparison with the Swanscombe Skull.* Some of the characters in the London skull fragment can also be compared with those of the prehistoric skull recently discovered at Swanscombe in Kent by Mr A. T. Marston. The Swanscombe skull is now admitted by all the leading authorities to have been found in the middle gravels of the 100 ft. terrace of the Thames and to be of Acheulean date. It consists of the complete occipital and left parietal bones, which articulate with one another very accurately. The parts preserved are much the same as in the London skull, except that in the latter the occipital and left parietal are not quite complete and a small portion of the right parietal is present. The bones in the Swanscombe skull, like those of the Piltdown skull, are much thicker than the cranial wall in the London skull, in which, as already mentioned, the thickness is much the same as in the average modern female skull. But both parietal and occipital bones as thick as those of the Swanscombe specimen are occasionally found in modern skulls with no evidence of disease.

The measurements of the Swanscombe specimen that can be compared with those of the London fragment are shown in Table III. In Figs. 5, 6 and 7 are also shown superimpositions of the dioptographic contour tracings of the London and Swanscombe skulls (the latter from a cast) in three planes, the sagittal, the horizontal and the transverse. In the profile contours the asterions and the lines indicating an approximation to the subcerebral plane, as determined by the direction of the left parieto-mastoid suture, have been made coincident. In the horizontal tracing the skulls are oriented in the subcerebral plane and the bregmatic points have been made to coincide. In the transverse maximum contour tracing the skulls are oriented at right angles to the subcerebral plane and the midpoints of the biasterionic diameters are coincident.

The maximum parietal breadth (B) in the Swanscombe skull (obtained by assuming the transverse contour of the right side to be the mirror image of that of the left at its widest part) is much the same as in the London skull, the respective measurements being 142 and 144 mm. In biasterionic breadth (B''') there is, however, an appreciable difference, the Swanscombe specimen being, like the Piltdown, wider in this region by about 7–9 mm. In this feature the Swanscombe skull, like the London (p. 304), probably exceeds the range of values found in skulls of modern type and of much the same size in their other principal dimensions. These differences are well illustrated in Fig. 7. The basio-bregmatic height (H') is not determinable accurately in the London skull, but in so far as it can be estimated from the other characters as approximately 120 mm. it seems to be definitely less than the corresponding height of the Swanscombe, which is 126 mm. (see Fig. 5). The latter is rather in excess of the average basio-

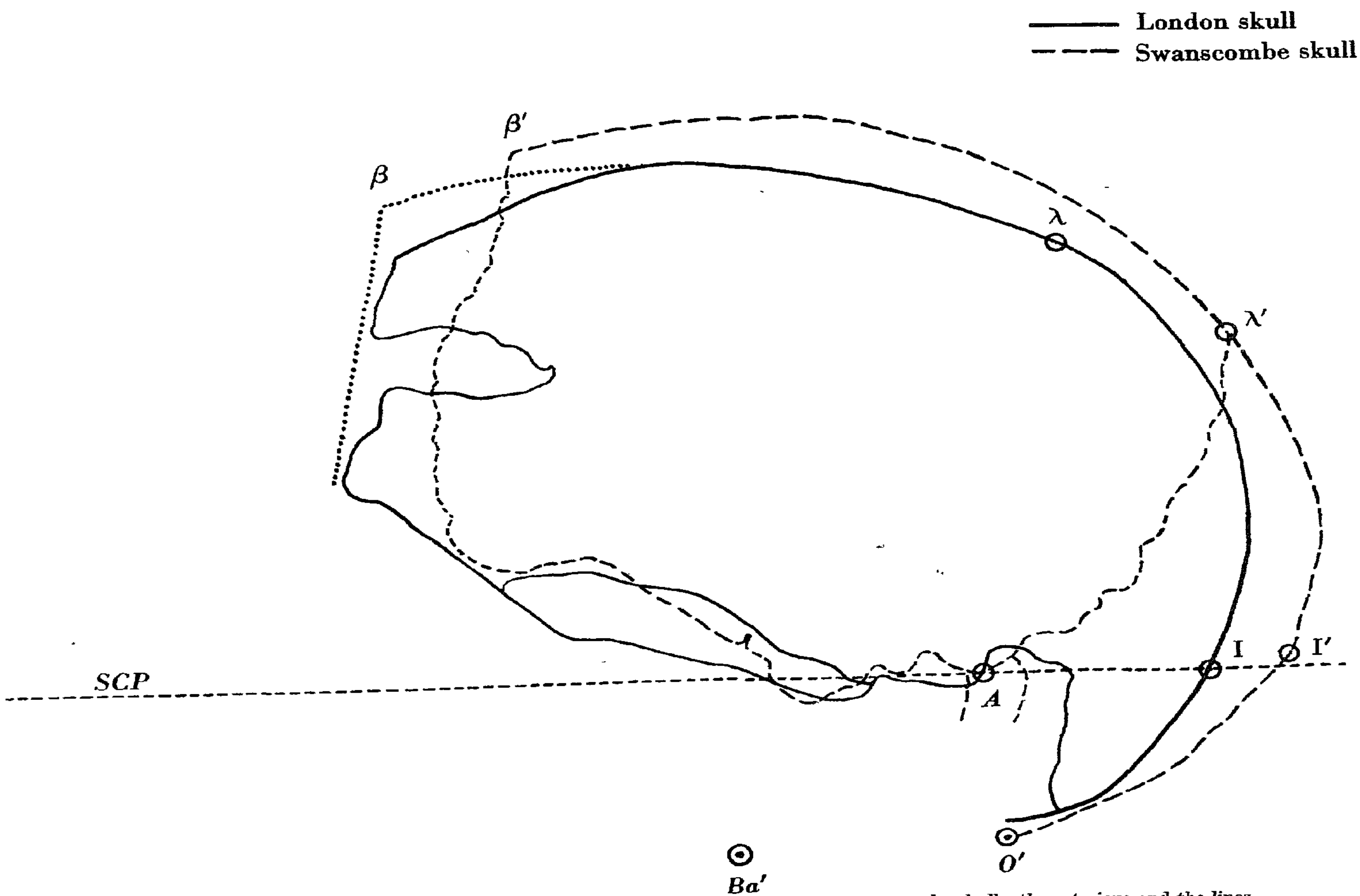


Fig. 5 Showing the superimposition of the sagittal tracings of the London and Swanscombe skulls, the asterions and the lines indicating the subcerebral plane being made coincident.

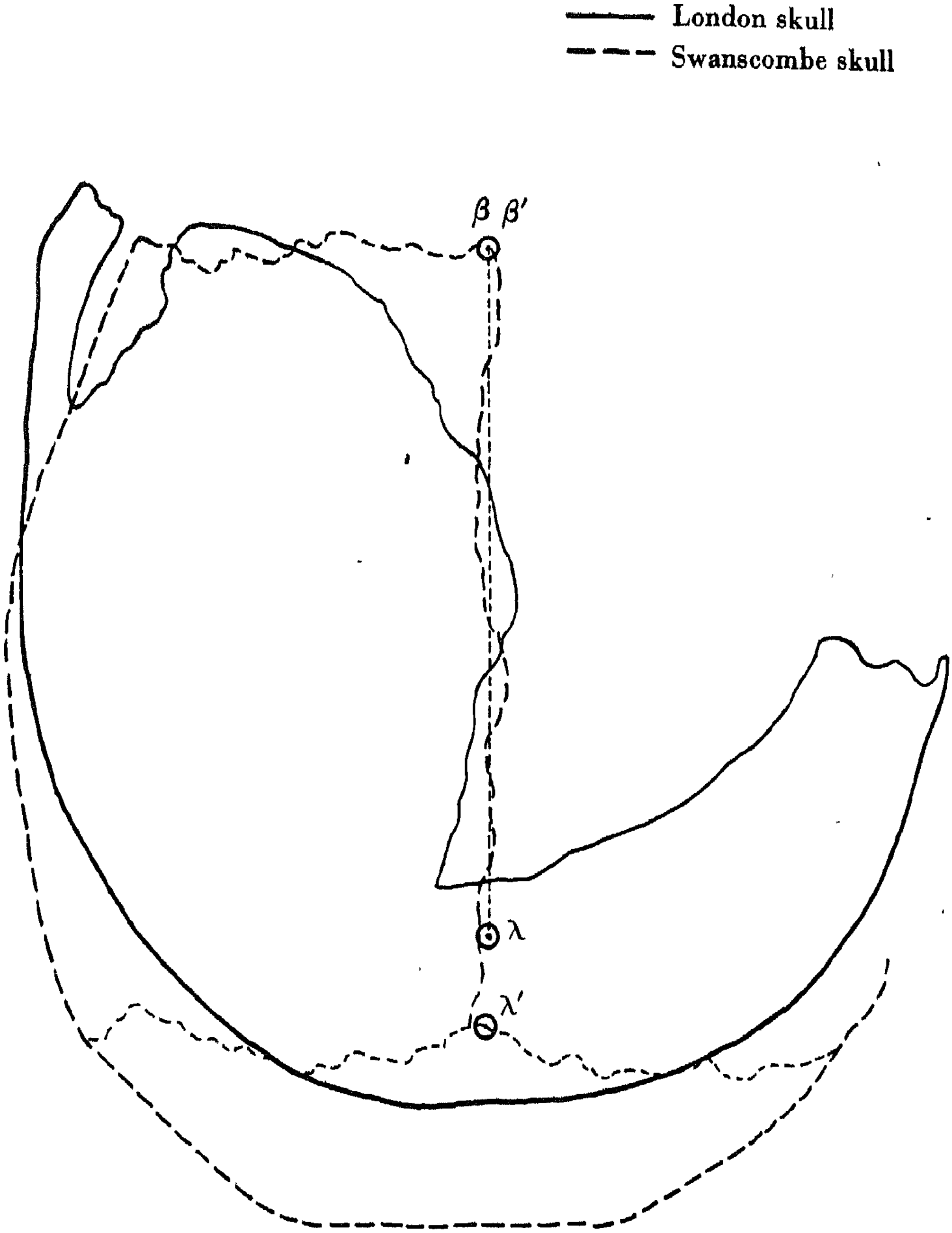


Fig. 6. Showing the superimposition of the maximum horizontal tracings of the London and Swanscombe skulls when the skulls are oriented in the subcerebral plane, and the bregmatic points are made coincident.

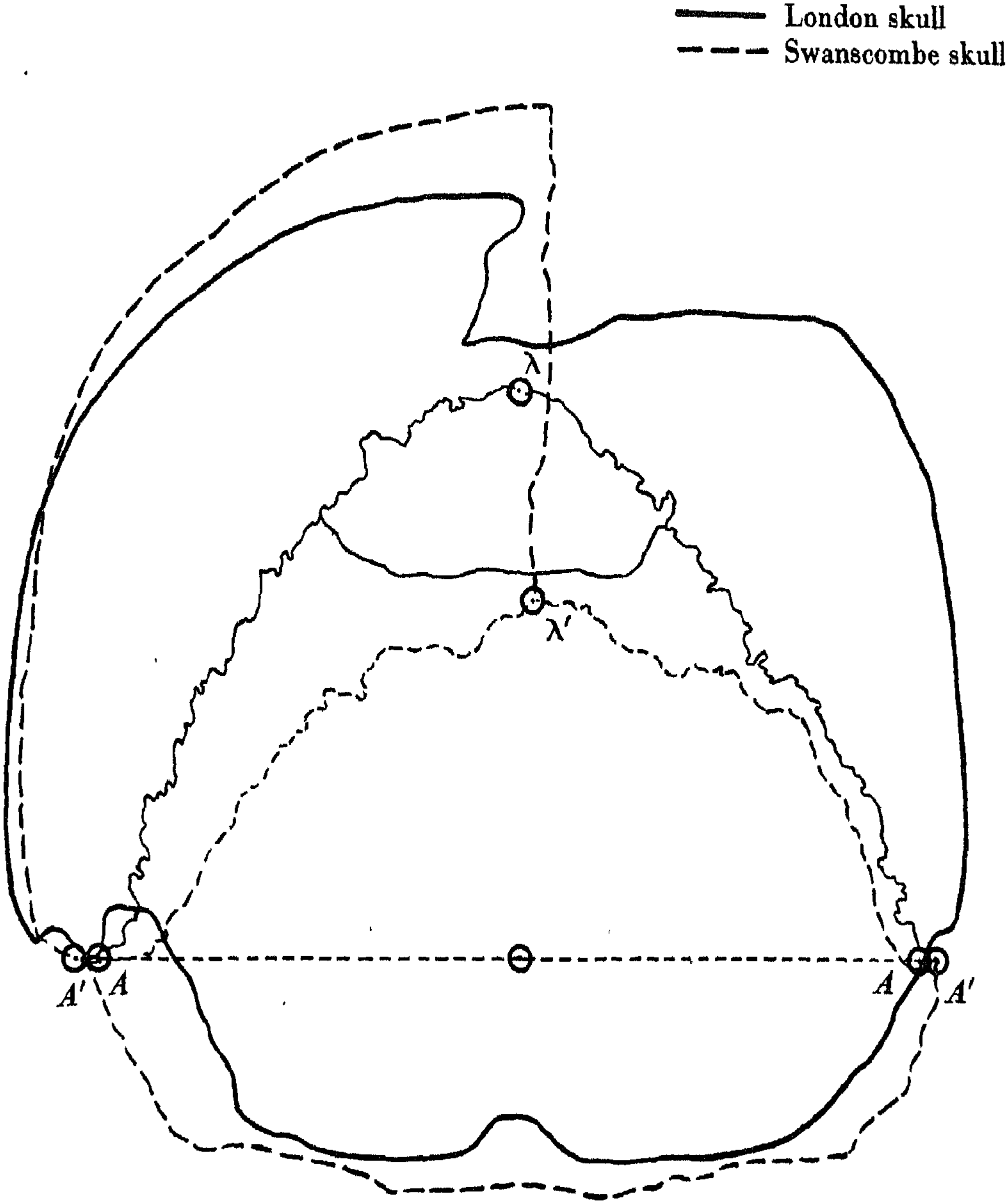


Fig. 7. Showing the superimposition of the maximum transverse tracings of the London and Swanscombe skulls when the skulls are oriented at right angles to the subcerebral plane and the mid-points of the biasterionic diameters are made coincident.

bregmatic height in the Farringdon Street female seventeenth-century Londoners (Table II) and coincides with the average for the modern West Scottish female. The maximum length (L) is not determinable accurately in the Swanscombe skull, but the length that results from the completion of the frontal region of the sagittal contour of the skull in the manner that seems not improbable from the segment that is extant is approximately 180 mm. This is much the same as the estimated maximum length for the London skull. The cephalic indices in the two skulls are thus probably in close agreement, 80 in the London and 79 in the Swanscombe. It is of interest to note that the maximum width of the Saccopastore skull is the same as in the Swanscombe skull and the maximum length is 181 mm., i.e. nearly the same, giving a cephalic index of 78·5 as compared with 79·0. The incomplete horizontal contour tracing of the Swanscombe skull suggests that when complete the specimen would have shown some post-orbital constriction, a feature that does not appear to be so clearly indicated in the London skull (Fig. 6).

Turning to the sagittal contours, the parietal arc in the Swanscombe is longer, and the occipital arc shorter, than in the London skull, but the lengths of these two arcs combined do not differ in the two skulls by more than 1 mm. The parietal arc is not quite so flat in the Swanscombe skull as in the London skull, as is shown by the respective values of the index ($100\ S_2'/S_2$) of 96·2 and 93·0. The occipital arc, though shorter in the Swanscombe specimen, has much the same curvature as in the London skull, the indices of occipital curvature ($100 \times S_3'/S_3$) being 81·7 and 81·3 and the values of Pearson's occipital index being 58·4 and 58·1. This similarity in curvature is also brought out by the fairly close agreement of the occipital (lambda-inion-opisthion) angles in the two skulls, 114° and 117°, and is shown in the contour tracings in Fig. 5. It is interesting to note that though both the occipital arc and the parietal arc of the Swanscombe skull are shorter than in the Piltdown, yet the curvatures of the corresponding bones are almost identical in the two specimens. In the superimposed sagittal contour tracings (Fig. 5), the projected line of the coronal suture in the Swanscombe skull seems to be approximately parallel to the direction in which the corresponding suture in the London skull is assumed to run, though when oriented as shown in the subcerebral plane, the antero-posterior axis of the foramen magnum in the former appears to be slightly tilted backward from the horizontal.

The supra-inial segment of the occipital arc is shorter and the infra-inial segment relatively longer in the Swanscombe than in the London skull. These differences may be related to the presence of the pre-interparietal bone in the London specimen, but the indices of curvature in the corresponding segments in the two skulls are of much the same order, as shown by the values 92·1 and 92·7, and 98·1 and 97·6. The proportion which the lambda-inion arc forms of the total occipital arc is less in the Swanscombe than in the London skull, 55 per cent. as

compared with 67 per cent. This ratio in the Swanscombe skull is in close agreement with that in the Piltdown.

As might be expected from the difference in biasterionic breadth, the ratio of the biasterionic breadth to the maximum parietal breadth is appreciably greater in the Swanscombe skull than in the London, the values being 85 and 79. With a greater biasterionic breadth and a shorter occipital chord in the Swanscombe the ratio 100 × biasterionic breadth/occipital arc in this specimen is also greater than in the London. On the other hand, the biasterionic breadth and the parietal chord in the Swanscombe exceed the corresponding measurements in the London specimen in much the same proportion, so that the ratio 100 × biasterionic breadth/parietal chord almost coincides in the two skulls, the indices being 114 and 113. This ratio in the Piltdown skull is also in close agreement with that in the Swanscombe.

So far as a study of the comparable measurements and their proportions in the Swanscombe and London skull fragments permits any inferences to be drawn, there would appear to be no unequivocal evidence that the Swanscombe skull shows any greater divergence from the modern type than does the London specimen.

6. *Summary and Conclusions.* The main inferences that can be drawn from the detailed comparisons that have been made may be summarized briefly:

1. From an anatomical point of view the London skull fragment apparently possesses no features other than the presence of a pre-interparietal bone which would be at all exceptional if found in a specimen of modern type.

2. Comparison of the fragment with the female skulls of well-authenticated Aurignacian and Solutrean date from Europe—considered as a group and individually—reveal the close resemblance between it and the skulls from Solutré. So far as can be judged from the available measurements, the London skull would appear to be indistinguishable in type from the skull usually designated Solutré V (1924).

3. Amongst the British skulls or fragments of skulls of reputed late palaeolithic or earlier date, excluding for the moment the Piltdown and Swanscombe skulls, the specimen that seems to suggest most strikingly a similarity of type with the London skull is the fragment of reputed late Acheulean date discovered in the vicinity of Bury St Edmunds. The most reasonable prediction of the missing parts of the vault in one fragment is in such close agreement with what is extant in the other that, while it is possible that the similarity may merely be a coincidence, it seems rather to support the view that the two fragments represent calvariae that were essentially similar in their proportions and general form.

4. From the differences that are observed in the mean measurements of two modern Scottish female groups—one (the normal) without and the other (the

anomalous) with interparietals—there seems to be a reasonable basis for the inference that the presence of the pre-interparietal bone in the London skull is probably largely responsible for the rather unusual formation of its occipital arc, in respect of extent and curvature, as well as for the relative shortness of its parietal arc.

5. Of the differences that are observed in the measurements of the characters in the London skull and the corresponding mean values in the Scottish series and in the closely related series of recent Londoners, only a few are of such a magnitude as to indicate that the measurements in the palaeolithic specimen may reasonably be considered exceptional for these series. Amongst these, the principal are the relatively long supra-inial and short infra-inial segments of the occipital arc and the relatively short and flat parietal arc. A disproportion in the segmental lengths of the occipital arc is not a primitive feature, as their relative proportions in the Gibraltar and the Piltdown skulls are identical with that found in the Scottish group with interparietals. It is possible that the relatively short infra-inial arc in the London skull may be to some extent compensatory to the long supra-inial arc.

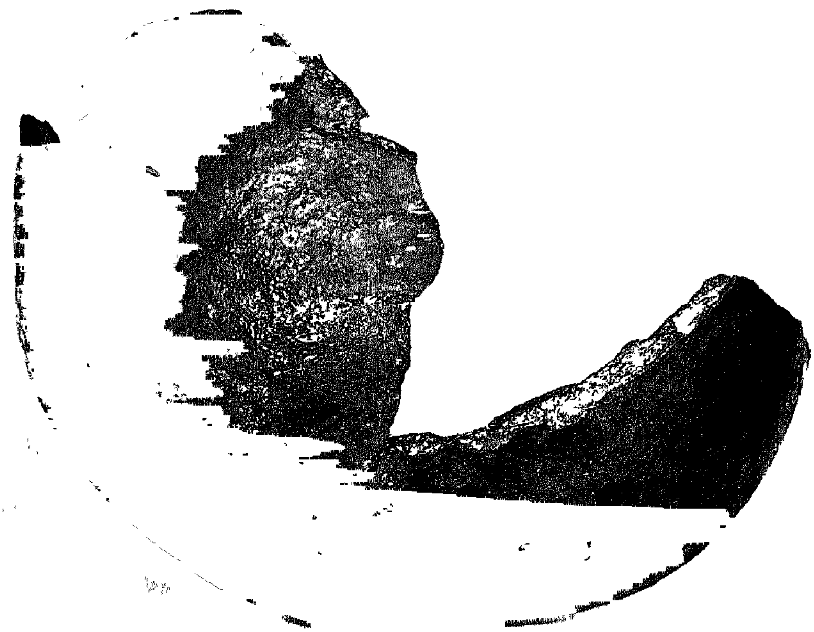
6. Comparison of the available measurements and their proportions in the London fragment with those of the corresponding characters in the Neanderthal female skulls from Gibraltar, La Quina and Saccopastore seems to indicate that, in its general form and dimensions, it resembles the last-named specimen most closely. In the unusual flatness of the parietal segment of the sagittal arc, it presents a definite Neanderthaloid feature, and possibly also in its relatively great biasterionic breadth, though the latter measurement is of the order occasionally found in modern types of skull.

7. In its measurements and proportions there is no unequivocal evidence that the London skull has a closer kinship with the Piltdown skull than with the modern Scottish female type. In the extent and curvature of its parietal arc, as well as in the proportions of biasterionic breadth to maximum parietal breadth and of biasterionic breadth to length of occipital chord, it is nearer to the Scottish than the Piltdown type, while in fullness of curvature of the lower occipital region it is in quite as close agreement with the one type as the other.

8. So far as can be ascertained from the comparison of measurements of the available characters and their proportions, there is no evidence that the Swanscombe skull shows a greater divergence from the modern type than does the London specimen.

It may be stated, finally, that the present study seems to indicate conclusively that the London skull is of the modern type and resembles closely in its general form that of the upper palaeolithic period found at Solutré, and especially the specimen designated Solutré V (1924). So far as can be judged from the fragments

A The London skull from the lateral aspect

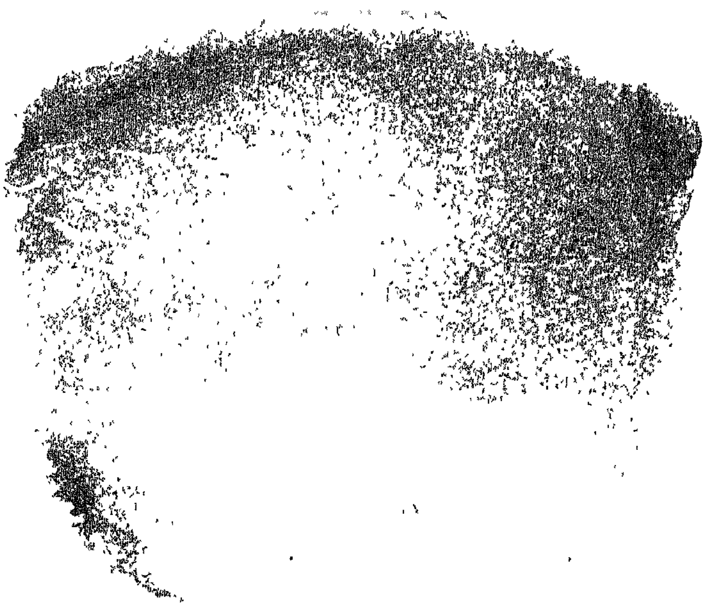


B. The London skull from the vertical aspect.

A. The London skull from the occipital aspect.

B The London skull from the internal aspect.

A. Radiograph of parietal bone of the London skull



B Radiograph of parietal bone of seventeenth-century Londoner.

that are preserved, it is also remarkably like the skull of the same presumptive age—i.e. late Acheulean or pre-Mousterian—found at Bury St Edmunds, Suffolk, and sometimes referred to as the Westley skull.

DESCRIPTION OF PLATES

Plate I. A. Showing the London skull from the lateral aspect when oriented in the subcerebral plane.

B. Showing the London skull from the vertical aspect when oriented in the subcerebral plane.

Plate II. A. Showing the London skull from the occipital aspect when oriented in the subcerebral plane. The line of fusion of the pre-interparietal bone is seen.

B. Showing the interior of the London skull viewed from the front (from a drawing by Mr A. K. Maxwell). *L*, crista lunata; *F*, deep fossa corticis striatae of the right side; *S*, small fossa corticis striatae of the left side.

Plate III. Showing a comparison of the radiograph of the parietal bone of the London skull (A) with that of a seventeenth-century Londoner (B) of much the same thickness. The high degree of mineralization of the former is clearly indicated. The densities of the two bones should be truly comparable, as the pictures were taken on the same film and print.

SIGNIFICANCE TESTS WHICH MAY BE APPLIED TO SAMPLES FROM ANY POPULATIONS

III.* THE ANALYSIS OF VARIANCE TEST

By E. J. G. PITMAN
University of Tasmania

1. *The two forms of the analysis of variance test.* The main title of this paper is, perhaps, not strictly accurate, for, in that form of the analysis of variance test discussed here, the observed numbers are not really regarded as a sample from a larger population, though, in an actual application of the test, the classification determined by "treatment" is only one of many possible. The essential point is that no assumptions are made about forms of populations. This, and the methods employed, link the paper naturally with the preceding papers (1, 2) of this series.

As one of a series, this paper was planned many months ago; but it was not written until June of this year, 1937. It arrived in England just about the time when B. L. Welch's paper (5) on the analysis of variance appeared. While some of its results are anticipated by Welch, the present paper goes deeper into the randomization theory of the simplest type of analysis of variance test.

The principles of this test may be briefly summarized as follows. Several batches, each consisting of n individuals, are taken and the individuals of a batch subjected to n different treatments, the allocation of the treatments to the individuals of a batch being determined by chance. Each individual is then measured, and we wish to determine whether the differences in treatment have produced any real differences in the character measured. The batches might, for example, be the blocks in an agricultural experiment, and the individuals the plots into which each block is subdivided.

If there are m batches, our observations consist of m sets of numbers,

$$
\begin{gathered}
a_1, a_2, \ldots, a_n, \\
b_1, b_2, \ldots, b_n, \\
\ldots\ldots\ldots\ldots,
\end{gathered}
$$

where $a_r, b_r, \ldots$ are the results for the m individuals subjected to treatment r. We assume that

$$a_r = A + T_r + x_{a_r}, \quad b_r = B + T_r + x_{b_r}, \quad \text{etc.},$$

where A denotes the result of some cause which affects equally all the individuals of the first batch, T_r denotes the effect of treatment r, and the third term x_{a_r} arises from the variability among the individuals of a batch, errors in measure-

* For the previous papers of this series see (1) and (2) in list of references.

ment, and other accidents affecting particular individuals. The analysis of variance is

$$S = S_B + S_T + S_E,$$

where S denotes the total sum of squares,

S_B is the sum of squares due to batches, and is independent of $T_1, T_2, \ldots,$

S_T is the sum of squares due to treatments, and is independent of $A, B, \ldots,$

S_E is the residual sum of squares, and is independent of both $A, B, \ldots,$ and $T_1, T_2, \ldots.$

Differences in the values of the T tend to increase the value of S_T while not affecting the value of S_E; hence large values of S_T/S_E are regarded as significant. The question then is, does the observed value of S_T/S_E indicate that the T are not all equal, or is it such as might easily arise by chance when the T are all equal?

It can be shown that if the x are independent chance variables each with the same normal distribution of standard deviation σ, and if $T_1 = T_2 = \ldots,$ then S_T/σ^2 and S_E/σ^2 are independent chance variables distributed like χ^2 with degrees of freedom $n-1$ and $(m-1)(n-1)$ respectively. Hence

$$W = \frac{S_T}{S_T + S_E} = \frac{S_T}{S - S_B},$$

which is a monotonic increasing function of S_T/S_E, has a

$$B\{\tfrac{1}{2}(n-1), \tfrac{1}{2}(m-1)(n-1)\}$$

distribution. This gives the usual test based on the above assumptions, though, in practice, some monotonic function of W such as Fisher's z, which is

$$\tfrac{1}{2}\log_e\{(m-1)W/(1-W)\},$$

is often employed. It should be noted that the theoretical repetitions which determine this distribution of W are repetitions of the whole experiment, and that the x values will be different samples from the same normal population.

The problem of testing the null hypothesis, that the T are all equal, has been tackled without making any assumptions about the x. If the null hypothesis is true, the observed value of W is the result of the chance allocation of the different treatments to the different individuals in the batches. We may imagine repetitions of the same experiment with the same batches and the same individuals, each with its corresponding x unaltered, but with different allocations of the treatments to the individuals in the various batches. If the T are all equal, the observed numbers $a_1, a_2, \ldots, b_1, b_2, \ldots$ will remain the same but will be arranged in different orders. There are $(n!)^{m-1}$ ways in which the numbers may be grouped into n groups each containing one and only one number from each batch, so that W may take $(n!)^{m-1}$ values, some of which may happen to coincide with one another. As the allocation of treatments to individuals is determined by chance,

all such groupings are equally likely, and so, therefore, are the corresponding values of W. To test the null hypothesis in this way, we must know this distribution of W, which is entirely determined by the observed numbers a_1, a_2, etc. We can then determine the probability that, if the null hypothesis were true, a value of W as great as, or greater than, that observed would be obtained. If this probability is small, say less than 0·05, we consider that the observed value of W is significant and that the differences in treatment have produced real differences in effect.

It is frequently stated that, for cases which commonly occur in certain fields of work, this distribution of W is approximately the same as the other distribution of W, that is, that this distribution of W is approximately a

$$B\left\{\tfrac{1}{2}(n-1),\ \tfrac{1}{2}(m-1)(n-1)\right\}$$

distribution. Eden and Yates(3), by a sampling process, showed that there was very good agreement between the B-distribution and the actual distribution of W in a case which they investigated. In order to discuss the question we shall obtain the first four moments of the exact distribution of W. Welch, in the paper(5) referred to above, obtained expressions for the first two moments.

2. *The moments of W.* Since

$$W=\frac{S_T}{S_T+S_E}=\frac{S_T}{S-S_B}$$

is independent of the quantities A, B, etc., we may assume these to have such values that the mean of each batch is zero. The mean of the whole set of mn numbers is then zero, and also $S_B=0$. We then have

$$W=\frac{S_T}{S}=\frac{\frac{1}{m}\sum\limits_{r=1}^{n}(a_r+b_r+\dots)^2}{\sum\limits_{r=1}^{n}a_r^2+\sum\limits_{r=1}^{n}b_r^2+\dots}.$$

The numbers $$a_1,\ a_2,\ \dots,\ a_n$$

are always a permutation of the same set of n numbers; the different values of W are obtained from the different permutations of the numbers in each batch. We shall denote the second, third, and fourth moments of the a by α_2, α_3, α_4, and their second, third, and fourth k-statistics (see (4), p. 75) by α_2', α_3', α_4'. We shall write

$$R_{ab}=a_1b_1+a_2b_2+\dots+a_nb_n.$$

There are ${}_mC_2=M$, say, of these expressions, which it will sometimes be convenient to denote by

$$R_1,\ R_2,\ \dots,\ R_M.$$

Putting $$U=R_1+R_2+\dots+R_M,$$

we have $$\Sigma\,(a_r+b_r+\dots)^2=\Sigma a_r^2+\Sigma b_r^2+\dots+2U,$$

and therefore
$$W=\frac{1}{m}+\frac{2U}{mn\Sigma\alpha_2}.$$

The moments of R_{ab} are as follows:

$$E\ (R_{ab})=0,$$

$$E\ (R_{ab}^2)=\frac{n^2\alpha_2\beta_2}{n-1},$$

$$E\ (R_{ab}^3)=\frac{n^3\alpha_3\beta_3}{(n-1)\ (n-2)}=\frac{(n-1)\ (n-2)\alpha_3'\beta_3'}{n},$$

$$E\ (R_{ab}^4)=\frac{3n^4\alpha_2^2\beta_2^2}{(n-1)\ (n+1)}+\frac{(n-1)\ (n-2)\ (n-3)\ \alpha_4'\beta_4'}{n\ (n+1)}.$$

These may be obtained directly without much labour; thus, for example,

$$\begin{aligned}E\ (R_{ab}^4)&=E\ \{(\Sigma a_p b_p)^4\}\\&=E\ \{\Sigma a_p^4 b_p^4+4\Sigma a_p^3 a_q b_p^3 b_q+6\Sigma a_p^2 a_q^2 b_p^2 b_q^2\\&\qquad+12\Sigma a_p^2 a_q a_r b_p^2 b_q b_r+24\Sigma a_p a_q a_r a_s b_p b_q b_r b_s\}\\&=nE\ (a_1^4)\ E\ (b_1^4)+4n\ (n-1)\ E\ (a_1^3 a_2)\ E\ (b_1^3 b_2)+\text{etc.}\end{aligned}$$

The mean value of such an expression as $a_1^3 a_2$ is easily obtained, for we have

$$0=\Sigma a_p\,\Sigma a_p^3=\Sigma a_p^3 a_q+\Sigma a_p^4=n\ (n-1)\ E\ (a_1^3 a_2)+n\alpha_4,$$

and therefore
$$E\ (a_1^3 a_2)=-\alpha_4/(n-1).$$

Proceeding in this way and then collecting terms, we obtain the result given above.

Before attempting to find the moments of

$$U=R_1+R_2+\ldots+R_M,$$

we must note that any two of the R are independent, and therefore

$$E\ (R_p R_q)=0.$$

Also, any three of the R are independent unless the three form a set like

$$R_{ab},\ R_{bc},\ R_{ca},$$

which (in the double-suffix notation) involves only three suffixes. Hence, in particular, for three R not related in this way,

$$E\ (R_p R_q R_r)=0.$$

In general, if in any product of any number of the R one suffix (in the double suffix notation) is not repeated, the corresponding R will be independent of the other R in the product, and therefore the mean value of the product will be zero.

This implies that the mean value of the product of four R will be zero unless the suffixes form a closed chain like

$$R_{ab},\ R_{bc},\ R_{cd},\ R_{da}.$$

$$E(U)=E(\Sigma R_p)=\Sigma E(R_p)=0.$$

$$E(U^2)=E(\Sigma R_p^2+2\Sigma R_p R_q)=\Sigma E(R_p^2)$$
$$=\frac{n^2}{n-1}\Sigma \alpha_2\beta_2.$$

$$E(U^3)=E(\Sigma R_p^3+3\Sigma R_p^2 R_q+6\Sigma R_p R_q R_r)$$
$$=\Sigma E(R_p^3)+6\Sigma E(R_{ab}R_{bc}R_{ca}).$$

Now $\quad E(R_{ab}R_{bc}R_{ca})=E(\Sigma a_p b_p\,.\,\Sigma b_p c_p\,.\,\Sigma c_p a_p)$

$$=E\{[\Sigma b_p^2 a_p c_p+\Sigma b_p b_q(a_p c_q+a_q c_p)]\,\Sigma c_p a_p\}$$
$$=E\{[E(b_1^2)\,\Sigma a_p c_p+E(b_1 b_2)\,\Sigma(a_p c_q+a_q c_p)]\,\Sigma c_p a_p\}.$$

But $\quad \Sigma a_p c_p+\Sigma(a_p c_q+a_q c_p)=\Sigma a_p\Sigma c_p=0,$

therefore $\quad \Sigma(a_p c_q+a_q c_p)=-\Sigma a_p c_p.$

Thus $\quad E(R_{ab}R_{bc}R_{ca})=\{E(b_1^2)-E(b_1 b_2)\}\,E\{(\Sigma a_p c_p)^2\}$

$$=\left(\beta_2+\frac{\beta_2}{n-1}\right)\frac{n^2\alpha_2\gamma_2}{n-1}=\frac{n^3\alpha_2\beta_2\gamma_2}{(n-1)^2}.$$

Since $\quad \Sigma E(R_p^3)=\dfrac{(n-1)(n-2)}{n}\Sigma\alpha_3'\beta_3',$

this gives $\quad E(U^3)=\dfrac{6n^3}{(n-1)^2}\Sigma\alpha_2\beta_2\gamma_2+\dfrac{(n-1)(n-2)}{n}\Sigma\alpha_3'\beta_3'.$

$$E(U^4)=E(\Sigma R_p^4+4\Sigma R_p^3 R_q+6\Sigma R_p^2 R_q^2+12\Sigma R_p^2 R_q R_r+24\Sigma R_p R_q R_r R_s)$$
$$=\Sigma E(R_p^4)+6\Sigma E(R_p^2)\,E(R_q^2)+12\Sigma E(R_{ab}^2 R_{bc} R_{ca})$$
$$+24\Sigma E(R_{ab}R_{bc}R_{cd}R_{da}).$$

$$\Sigma E(R_p^2)\,E(R_q^2)=\tfrac{1}{2}\{[\Sigma E(R_p^2)]^2-\Sigma[E(R_p^2)]^2\}$$
$$=\frac{n^4}{2(n-1)^2}\{[\Sigma\alpha_2\beta_2]^2-\Sigma\alpha_2^2\beta_2^2\}.$$

$$E(R_{ab}^2 R_{bc}R_{ca})=E\{(\Sigma a_p b_p)^2\,.\,\Sigma b_p c_p\,.\,\Sigma c_p a_p\}$$
$$=E\{(\Sigma a_p b_p)^2\,(\Sigma c_p^2 a_p b_p+\Sigma c_p c_q(a_p b_q+a_q b_p))\}$$
$$=E\{(\Sigma a_p b_p)^2\,(E(c_1^2)-E(c_1 c_2))\,\Sigma a_p b_p\}$$
$$=\{E(c_1^2)-E(c_1 c_2)\}\,E\{(\Sigma a_p b_p)^3\}$$
$$=\frac{n\gamma_2}{n-1}\,\frac{(n-1)(n-2)\,\alpha_3'\beta_3'}{n}$$
$$=(n-2)\,\alpha_3'\beta_3'\gamma_2.$$

$$\begin{aligned}E(R_{ab}R_{bc}R_{cd}R_{da}) &= E(\Sigma a_p b_p . \Sigma b_p c_p . \Sigma c_p d_p . \Sigma d_p a_p)\\ &= E\{(\Sigma b_p^2 a_p c_p + \Sigma b_p b_q (a_p c_q + a_q c_p))(\Sigma d_p^2 a_p c_p + \Sigma d_p d_q (a_p c_q + a_q c_p))\}\\ &= E\{(E(b_1^2) - E(b_1 b_2)) \Sigma a_p c_p . (E(d_1^2) - E(d_1 d_2)) \Sigma a_p c_p\}\\ &= \{E(b_1^2) - E(b_1 b_2)\}\{E(d_1^2) - E(d_1 d_2)\} E\{(\Sigma a_p c_p)^2\}\\ &= \frac{n^4}{(n-1)^3} \alpha_2 \beta_2 \gamma_2 \delta_2 .\end{aligned}$$

Since four letters can be arranged in a closed chain in three ways, this gives

$$E\{\Sigma R_{ab}R_{bc}R_{cd}R_{da}\} = \frac{3n^4}{(n-1)^3} \Sigma \alpha_2 \beta_2 \gamma_2 \delta_2 .$$

Thus
$$\begin{aligned}E(U^4) = &\frac{3n^4}{(n-1)(n+1)} \Sigma \alpha_2^2 \beta_2^2 + \frac{(n-1)(n-2)(n-3)}{n(n+1)} \Sigma \alpha_4' \beta_4'\\ &+ \frac{3n^4}{(n-1)^2} \{(\Sigma \alpha_2 \beta_2)^2 - \Sigma \alpha_2^2 \beta_2^2\}\\ &+ 12(n-2) \Sigma \alpha_3' \beta_3' \gamma_2 + \frac{3.24n^4}{(n-1)^3} \Sigma \alpha_2 \beta_2 \gamma_2 \delta_2 .\end{aligned}$$

Finally, for the moments of

$$W = \frac{1}{m} + \frac{2U}{mn\Sigma\alpha_2},$$

we have* $E(W) = \dfrac{1}{m}$

$$\begin{aligned}E\{(W-\overline{W})^2\} &= \frac{4}{m^2(n-1)} \frac{\Sigma \alpha_2 \beta_2}{(\Sigma \alpha_2)^2}\\ E\{(W-\overline{W})^3\} &= \frac{48}{m^3(n-1)^2} \frac{\Sigma \alpha_2 \beta_2 \gamma_2}{(\Sigma \alpha_2)^3} + \frac{8(n-1)(n-2)}{m^3 n^4} \frac{\Sigma \alpha_3' \beta_3'}{(\Sigma \alpha_2)^3}\\ E\{(W-\overline{W})^4\} &= \frac{48}{m^4(n-1)^2} \frac{(\Sigma \alpha_2 \beta_2)^2}{(\Sigma \alpha_2)^4} - \frac{96}{m^4(n-1)^2(n+1)} \frac{\Sigma \alpha_2^2 \beta_2^2}{(\Sigma \alpha_2)^4}\\ &\quad + \frac{72.16}{m^4(n-1)^3} \frac{\Sigma \alpha_2 \beta_2 \gamma_2 \delta_2}{(\Sigma \alpha_2)^4} + \frac{16(n-1)(n-2)(n-3)}{m^4 n^5 (n+1)} \frac{\Sigma \alpha_4' \beta_4'}{(\Sigma \alpha_2)^4}\\ &\quad + \frac{16.12(n-2)}{m^4 n^4} \frac{\Sigma \alpha_3' \beta_3' \gamma_2}{(\Sigma \alpha_2)^4} .\end{aligned}$$

3. *Comparison of the W- and B-distributions.* Only the first moment of W is independent of the particular numbers $a_1, a_2, \ldots, b_1, b_2, \ldots$. The mean and the variance of the $B\{\frac{1}{2}(n-1), \frac{1}{2}(m-1)(n-1)\}$ distribution are

$$\frac{1}{m} \text{ and } \frac{2(m-1)}{m^2(mn-m+2)}$$

respectively, so that W has always the correct mean value. Its range also is

* The expressions for $E(W)$ and $E\{(W-\overline{W})^2\}$ were given by Welch(5) but not those for the third and fourth moments. The expression K defined in section 3 is equal to Welch's $1-A$.

right, since W must lie between 0 and 1. But its variance is not necessarily correct. If W has approximately a $B\{\frac{1}{2}(n-1), \frac{1}{2}(m-1)(n-1)\}$ distribution, we must have approximately

$$\frac{4}{m^2(n-1)}\frac{\Sigma\alpha_2\beta_2}{(\Sigma\alpha_2)^2}=\frac{2(m-1)}{m^2(mn-m+2)},$$

that is

$$\frac{2\Sigma\alpha_2\beta_2}{(\Sigma\alpha_2)^2}=\frac{(m-1)(n-1)}{mn-m+2}.$$

Now

$$\frac{2\Sigma\alpha_2\beta_2}{(\Sigma\alpha_2)^2},$$

which we shall denote by K, may have any value between 0 and $(m-1)/m$. It approaches the lower limit when one of the quantities,

$$\alpha_2, \beta_2, \gamma_2, \ldots$$

becomes much larger than all the rest, and it takes the upper limit, $(m-1)/m$, as value when

$$\alpha_2=\beta_2=\gamma_2=\ldots.$$

Hence all that can be said in general about the variance of W is that it is not greater than

$$\frac{2(m-1)}{m^3(n-1)},$$

and that it takes this value when the variance of each batch is the same.

The Eden & Yates experiment (3) was equivalent to taking a sample of a thousand values of W derived from the sets

100	92	0	108
71	0	119	170
197	0	149	161
0	334	140	90
75	43	0	6
0	12	269	337
0	184	71	195
104	100	0	116

Their results showed very good agreement between the W distribution and the $B\{\frac{1}{2}(n-1), \frac{1}{2}(m-1)(n-1)\}$ distribution, in this case a $B\{1\frac{1}{2}, 10\frac{1}{2}\}$ distribution. For this to be so, the value of K must be approximately

$$\frac{(m-1)(n-1)}{mn-m+2}=\frac{21}{26}=0{\cdot}8077.$$

The batch variances multiplied by four are

7628,	15702,	22669,	59732,
3666,	90593,	26297,	8672,

from which we obtain $K=0{\cdot}7577$,

which is about $\frac{15}{16}$ of the required value. The moments of the two distributions are:

	$B(1\frac{1}{2}, 10\frac{1}{2})$	W
Mean value	0·125	0·125
Variance	0·008413	0·007893
Third moment about mean	0·000901	0·000733
Fourth moment about mean	0·000319	0·000246

From these we should expect that a sample of a thousand values of W would be well fitted by the B-distribution, for the differences between corresponding moments are rather small to be shown up by a sample of a thousand. The standard deviation of the variance of a sample of a thousand values of W is 0·00043, and the W and B variances differ by only about $1\frac{1}{5}$ times this. It is essentially the particular value of K which makes for good agreement.

Usually the terms involving α_3', etc., α_4', etc., will be negligible in comparison with the other terms in the expressions for the moments of W. Assuming that this is so, we have

$$0 \leqq W \leqq 1,$$

$$E(W) = \frac{1}{m},$$

$$E\{(W-\overline{W})^2\} = \frac{2}{m^2(n-1)} \frac{2\Sigma\alpha_2\beta_2}{(\Sigma\alpha_2)^2},$$

and approximately,

$$E\{(W-\overline{W})^3\} = \frac{48}{m^3(n-1)^2} \frac{\Sigma\alpha_2\beta_2\gamma_2}{(\Sigma\alpha_2)^3},$$

$$E\{(W-\overline{W})^4\} = \frac{12}{m^4(n-1)^2} \frac{(2\Sigma\alpha_2\beta_2)^2}{(\Sigma\alpha_2)^4} + \frac{72.16}{m^4(n-1)^3} \frac{\Sigma\alpha_2\beta_2\gamma_2\delta_2}{(\Sigma\alpha_2)^4} - \frac{96}{m^4(n-1)^2(n+1)} \frac{\Sigma\alpha_2^2\beta_2^2}{(\Sigma\alpha_2)^4}.$$

If the value of K is approximately

$$\frac{(m-1)(n-1)}{mn-m+2},$$

the distribution of W will be approximately a $B\{\frac{1}{2}(n-1), \frac{1}{2}(m-1)(n-1)\}$ distribution, for the range and the mean will be right, the variance will be approximately right, and it will generally be found that the third and fourth moments are approximately right. If a few of the batch variances are very much larger than the rest, the value of K will be too small. In this case there are three alternative procedures. We might discard these batches; if retained without modification they will dominate the experiment. If this is not desirable we might fit a B-distribution by use of the first two moments of W*. The third alternative

* This method has been investigated by Welch in the case of a few uniformity trials ((5) p. 31).

is to make all the batch variances equal by multiplying each batch of numbers by a suitable constant. There seems to be no theoretical objection to this as a preliminary to testing the null hypothesis, and it has the advantage that we then know a great deal more about the distribution of W; but it has the practical disadvantage of involving a lot of calculation. It is mentioned here merely as a theoretical possibility.

If the batch variances $\alpha_2, \beta_2, \ldots$ are all equal, or approximately equal, the value of K will be too large, but this will have an appreciable effect only when m and n are both fairly small, for when the batch variances are equal K takes its maximum value,

$$\frac{m-1}{m}, \qquad \ldots\ldots\text{(I)}$$

which, when $m(n-1)$ is large, is very close to

$$\frac{(m-1)(n-1)}{mn-m+2}=\frac{m-1}{m}\,\frac{1}{1+\dfrac{2}{m(n-1)}}, \qquad \ldots\ldots\text{(II)}$$

the value required for the $B\{\tfrac{1}{2}(n-1), \tfrac{1}{2}(m-1)(n-1)\}$ distribution. Moreover, K is fairly insensitive to changes in the values of the batch variances when m is large; inequalities in the batch variances will make K less than its maximum value (I) and therefore, provided they are not too great, fairly close to the required value (II).

The tables below show the values of (II) for small values of m and n, and the values of K for various values of the batch variances.

Values of $\dfrac{(m-1)(n-1)}{mn-m+2}$

	n						
m	3	4	5	6	7	8	∞
3	0·500	0·545	0·571	0·588	0·600	0·609	0·667
4	0·600	0·643	0·667	0·682	0·692	0·700	0·750
5	0·667	0·706	0·727	0·741	0·750	0·757	0·800
6	0·714	0·750	0·769	0·781	0·789	0·795	0·833

$m=3$

Batch variances	K
1, 1, 1	0·667
1, 1·5, 2	0·642
1, 2, 2	0·640
1, 1, 2	0·625

$m=3$

Batch variances	K
1, 2, 3	0·611
1, 2, 4	0·571
1, 1, 3	0·560
1, 1, 4	0·500

$m=4$

Batch variances	K
1, 1, 1, 1	0·750
1, 1·5, 2, 2·5	0·724
1, 1, 1, 2	0·720
1, 2, 3, 4	0·700
1, 1, 4, 4	0·660
1, 2, 4, 8	0·622

$m=5$

Batch variances	K
1, 1, 1, 1, 1	0·800
1, 3, 3, 3, 5	0·764
1, 2, 3, 4, 5	0·756
1, 1, 4, 6, 6	0·722
1, 1, 1, 4, 4	0·711
1, 1, 2, 4, 8	0·664

Since, in applying the significance test, we require only a rough approximation to the probability of obtaining, if the null hypothesis were true, a value of W not less than that observed, the $B\{\frac{1}{2}(n-1), \frac{1}{2}(m-1)(n-1)\}$ distribution will very frequently be a sufficiently good approximation to the distribution of W, especially when m and n are large. When either m or n is less than 5, the value of K should be calculated. If this differs considerably from

$$(m-1)(n-1)/(mn-m+2),$$

we can, as suggested above, either fit a B-distribution to the distribution of W by means of the first two moments of W, or equalize the batch variances and use the B-distribution discussed in the next section.

We shall now show that if a B-distribution is fitted by means of the first two moments of W, the third and fourth moments will agree well provided that K is not too small, and hence we may expect a good fit. In other words, if K is not too small, the distribution of W is approximately a B-distribution with mean $1/m$ and variance $2K/\{m^2(n-1)\}$.

From the relation

$$\begin{aligned} 3\Sigma\alpha_2\beta_2\gamma_2 &= \Sigma\alpha_2 . \Sigma\alpha_2\beta_2 - \Sigma\alpha_2^2\beta_2 \\ &= \Sigma\alpha_2 . \Sigma\alpha_2\beta_2 - (\Sigma\alpha_2 . \Sigma\alpha_2^2 - \Sigma\alpha_2^3) \\ &= \Sigma\alpha_2\{3\Sigma\alpha_2\beta_2 - (\Sigma\alpha_2)^2\} + \Sigma\alpha_2^3, \end{aligned}$$

we have
$$\frac{6\Sigma\alpha_2\beta_2\gamma_2}{(\Sigma\alpha_2)^3} = 3K - 2 + \frac{2\Sigma\alpha_2^3}{(\Sigma\alpha_2)^3}.$$

Since the α_2 are all positive,

$$\Sigma\alpha_2\Sigma\alpha_2^3 \geqq (\Sigma\alpha_2^2)^2,$$

and therefore
$$\frac{\Sigma\alpha_2^3}{(\Sigma\alpha_2)^3} \geqq \left\{\frac{\Sigma\alpha_2^2}{(\Sigma\alpha_2)^2}\right\}^2 = (1-K)^2.$$

Thus
$$\frac{6\Sigma\alpha_2\beta_2\gamma_2}{(\Sigma\alpha_2)^3} \geqq 3K - 2 + 2(1-K)^2 = K^2\left(1 - \frac{1-K}{K}\right).$$

Again, $$(\Sigma\alpha_2^2)^3 = \Sigma\alpha_2^2\,(\Sigma\alpha_2^2)^2 > \Sigma\alpha_2^2\Sigma\alpha_2^4 \geqq (\Sigma\alpha_2^3)^2;$$

therefore $$\frac{\Sigma\alpha_2^3}{(\Sigma\alpha_2)^3} < \left\{\frac{\Sigma\alpha_2^2}{(\Sigma\alpha_2)^2}\right\}^{\frac{3}{2}} = (1-K)^{\frac{3}{2}} = (1-K)\,(1-K)^{\frac{1}{2}}$$
$$< (1-K)\,(1-\tfrac{1}{2}K-\tfrac{1}{8}K^2),$$

that is $$< 1-\tfrac{3}{2}K+\tfrac{3}{8}K^2+\tfrac{1}{8}K^3.$$

Hence $$\frac{6\Sigma\alpha_2\beta_2\gamma_2}{(\Sigma\alpha_2)^3} < K^2\frac{3+K}{4}.$$

The third moment about the mean of the B-distribution which has the same first two moments as W is given by

$$\bar{\mu}_3 = \frac{8K^2}{m^3\,(n-1)^2}\,\frac{m-2}{m-1+2K/(n-1)}.$$

The third moment about the mean of the W-distribution is approximately

$$\frac{8K^2}{m^3\,(n-1)^2}\,\frac{6\Sigma\alpha_2\beta_2\gamma_2}{K^2\,(\Sigma\alpha_2)^3}.$$

Since $$\frac{6\Sigma\alpha_2\beta_2\gamma_2}{K^2\,(\Sigma\alpha_2)^3}$$

lies between $$1-\frac{1-K}{K} \text{ and } \frac{3+K}{4},$$

the third moments will be approximately the same provided that K is not too small; for example, with $K=\frac{3}{4}$, $E\,(W-\overline{W})^3$ lies between

$$\frac{2}{3}\,\frac{8K^2}{m^3\,(n-1)^2} \text{ and } \frac{15}{16}\,\frac{8K^2}{m^3\,(n-1)^2}.$$

If $m=5$, $n=4$, the value of $\bar{\mu}_3$ is

$$\frac{2}{3}\,\frac{8K^2}{m^3\,(n-1)^2}.$$

It should be noted that if $K=\frac{3}{4}$, m cannot be less than 4, and that if $m=4$ the batch variances are all the same. Hence

$$E\,(W-\overline{W})^3 = \frac{2}{3}\,\frac{8K^2}{m^3\,(n-1)^2},\quad \bar{\mu}_3 = \frac{2}{3+3/(2n-2)}\,\frac{8K^2}{m^3\,(n-1)^2}.$$

The fourth moment about the mean of the B-distribution is

$$\frac{12K^2}{m^4\,(n-1)^2}+\frac{48K^3}{m^4\,(n-1)^3},$$

neglecting terms of higher order in $1/m$ and $1/(n-1)$. The first term is the same as the first term in the expression for $E\,(W-\overline{W})^4$. The second term in the expression for $E\,(W-\overline{W})^4$ is

$$\frac{48}{m^4\,(n-1)^3}\,\frac{24\Sigma\alpha_2\beta_2\gamma_2\delta_2}{(\Sigma\alpha_2)^4},$$

which is positive but less than

$$\frac{48}{m^4 (n-1)^3} K^2 \frac{3+K}{4},$$

since $$24\Sigma\alpha_2\beta_2\gamma_2\delta_2 < \Sigma\alpha_2 . 6\Sigma\alpha_2\beta_2\gamma_2 < (\Sigma\alpha_2)^4 K^2 \frac{3+K}{4}.$$

Unless K is very small the third term in the expression for $E(W-\overline{W})^4$ will be negligible in comparison with these, and hence the fourth moments will be about the same.

Without definite knowledge about the batch variances, we cannot discuss further the values of the third and fourth moments of W. It is therefore of some interest to consider the values of these moments for the particular case of equal batch variances, and to see how they agree with the corresponding moments of a B-distribution which has the same first two moments as W. This is a limiting case, and will give us some idea of the truth in other cases, especially when m and n are large. In view of what has already been said this particular case may not be without practical importance. Further, the distribution of W when the batch variances are all equal has a direct practical application in cases where the individuals in a batch are not measured but merely graded or ranked with respect to some character. Our observations then consist of m sets of numbers, each set a permutation of the integers from 1 to n. If the different treatments really produce different effects on the character by which the individuals are ranked, the value of W will tend to be large, for individuals subjected to the same treatment will tend to have about the same rank in each batch. Large values of W are therefore significant. To test for significance, we must know the distribution of W when treatments are ineffective, that is when all possible associations of the rank numbers from the various batches are equally probable.

4. *The approximate distribution of W when the batch variances are equal.*

If $$\alpha_2 = \beta_2 = \gamma_2 = \ldots,$$

we have $$E(W) = \frac{1}{m},$$

$$E\{(W-\overline{W})^2\} = \frac{2(m-1)}{m^3(n-1)}.$$

Assuming, as before, that the terms involving α_3', etc., α_4', etc. are negligible, we have approximately

$$E\{(W-\overline{W})^3\} = \frac{8(m-1)(m-2)}{m^5(n-1)^2},$$

$$E\{(W-\overline{W})^4\} = \frac{12(m-1)^2}{m^6(n-1)^2} + \frac{48(m-1)(m-2)(m-3)}{m^7(n-1)^3} - \frac{48(m-1)}{m^7(n-1)^2(n+1)}.$$

The $B(p, q)$ distribution has the same first two moments as W if

$$p = \frac{n-1}{2} - \frac{1}{m},$$

$$q = \frac{(m-1)(n-1)}{2} - \frac{m-1}{m}.$$

The third moment about the mean of this $B(p, q)$ distribution is

$$\frac{8(m-1)(m-2)}{m^4(n-1)(mn-m+2)} = \frac{8(m-1)(m-2)}{m^5(n-1)^2}\left\{1 - \frac{2}{m(n-1)+2}\right\},$$

which agrees well with the corresponding value above when $m(n-1)$ is not too small. The fourth moment about the mean of this $B(p, q)$ distribution is

$$\frac{12(m-1)}{m^6(n-1)^2}\,\frac{(m-1)y^2 + 4m^2y - 14(m-1)y}{(y+2)(y+4)},$$

where $$y = m(n-1).$$

The difference between this and the corresponding expression above will be found to be of the same order as the third term in the expression for $E\{(W-\overline{W})^4\}$ and therefore negligible if $m(n-1)$ is not too small. The $B(p, q)$ distribution will thus be a fairly good approximation provided that $m(n-1)$ is not too small.

In order to test the agreement for rather small values of m and n, the following sets of numbers were taken:

-6	-2	3	5
-5	-3	1	7
-3	0	1	4

Numbers proportional to these but with equal batch variances (1/36) are

$-0{\cdot}23250$	$-0{\cdot}07750$	0·11625	0·19375
$-0{\cdot}18185$	$-0{\cdot}10911$	0·03637	0·25459
$-0{\cdot}26726$	0	0·08909	0·17817

The 576 values of W were calculated, and the following table shows P, the true probability, and P', the probability calculated from the $B(1\tfrac{1}{6}, 2\tfrac{1}{3})$ distribution, of obtaining a value of W as great as, or greater than, that shown.

W	P	P'	W	P	P'
0·09	0·894	0·855	0 50	0·253	0·238
0·12	0·797	0·801	0·56	0·188	0·178
0·17	0·691	0·713	0·66	0·099	0·099
0·22	0·595	0·627	0·76	0·050	0 045
0·29	0·493	0·514	0·80	0·026	0·029
0·36	0·396	0·411	0·83	0 019	0·020
0 44	0·297	0·306	0·87	0 010	0·011

For such small values of m and n the agreement is fairly good. It is very good at the upper tail, which is what we are interested in when applying the significance test.

It should be noticed that when m and n are large this $B(p, q)$ distribution will differ very slightly from the $B\{\frac{1}{2}(n-1), \frac{1}{2}(m-1)(n-1)\}$ distribution.

Only the simplest type of analysis of variance test has been discussed in this paper. I had intended another paper to follow, which would deal in the same way with the Latin square arrangement; but this has been dealt with by Welch (5). I may add that Welch's equation (49) on p. 41, giving the variance of W for the Latin square, agrees with my own result, which was reached by a route quite different from his. In view of the rather heavy algebra involved it seems worth while publishing this confirmation of Welch's result.

SUMMARY

The form of the analysis of variance test which involves no assumptions of normality is discussed. Expressions for the first four moments of the statistic used in this test are obtained. From these it appears that when the number of individuals in each batch, and the number of batches are both not too small, the usual test may be safely applied. A method of testing the validity of the approximation which this test employs is stated, and modifications of procedure, when necessary, are suggested.

REFERENCES

(1) PITMAN, E. J. G. (1937). *J. Roy. Statist. Soc.* Suppl. Vol. 4, pp. 119–130.

(2) —— (1937). *J. Roy. Statist. Soc.* Suppl. Vol. 4, pp. 225–232.

(3) EDEN, T. & YATES, F. (1933). *J. Agric. Sci.* Vol. 23, pp. 6–16.

(4) FISHER, R. A. (1936). *Statistical Methods for Research Workers.* 6th Edition (Oliver and Boyd).

(5) WELCH, B. L (1937). *Biometrika*, Vol. 29, pp. 21–52.

ABSENCE AND RECOVERY RECORDED AT IRREGULAR INTERVALS, ILLUSTRATED ON SCHOOL DATA

By FRANK SANDON

CONTENTS

(1) The sickness recovery curve

In any population at risk there is a certain morbidity rate, and ideally there would be machinery for recording the number falling sick at any instant, or more precisely, per unit time. Of those who so fall sick, the great majority recover, and ideally that fraction so recovering of the number first recorded would be recorded also at any subsequent instant. If the death-rate from sickness is nil we should then have a sickness recovery curve (S.R.C.) which would approach asymptotically with unlimited time the value zero, the curve giving at any instant the number still sick out of those who fell sick at any specified previous instant.

We may further suppose that of the number who at any particular instant become unfit all are at once as sick then as they will be, i.e. there is no incubation period, and that they all immediately begin, in varying degrees and at varying rates, to recover. The S.R.C. will therefore be a J curve, monotone in its decrease with time. Thus in a particular day secondary school, taking boys and girls of ages 10–19, a table of duration of absence was as follows:

TABLE A

Duration in sessions*	1	2	3	4	5	6	7	8	9	10	11
Frequency	790	403	169	76	111	41	23	31	18	43	11
Duration in sessions	12	13	14	15	16	17	18	19	20	21	22
Frequency	8	13	5	15	5	4	8	1	9	3	0

In addition: 18 cases with duration 23–30 sessions. Total: 1805

* It will be seen below that the concept of "session" as a unit of time is unsatisfactory: this is probably in part the cause of the unusual "humpiness" of the curve.

As is usual with abrupt curves we cannot say whether the curve is very strongly skewed with a mode between 0 and 1 or whether it is a J curve. The

latter as above indicated seems a plausible assumption, and we shall therefore adopt it in the belief that we are not far out in so doing.

(2) THE SICKNESS RECOVERY SURFACE

In any population, however, we should not simply record at any instant the number still sick of those who fell sick at a specific previous instant. We should have the "survivors", i.e. those still sick, of those who fell sick at *all* previous times, and we must now consider this extension of the problem. We may assume as a first approximation that there is a definite rate, in the population, of falling sick, and that this is constant and independent of time. This assumes

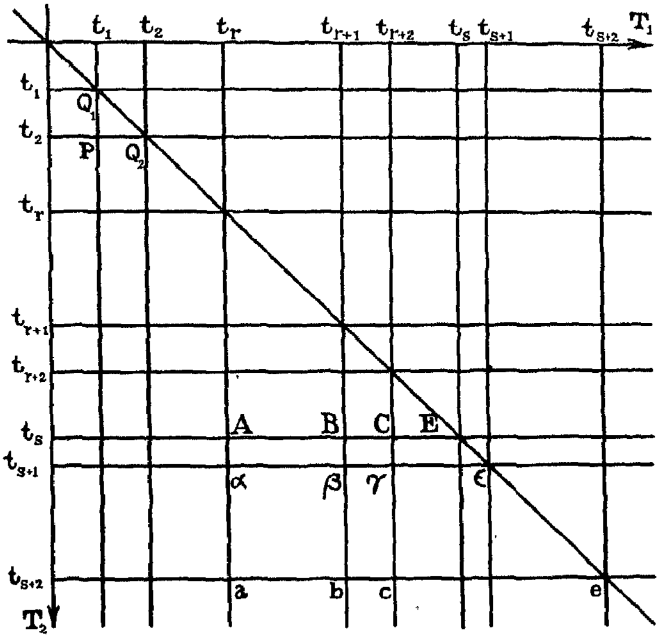


Fig. 1.

that there are no cyclical fluctuations nor epidemics, nor even random variations, a state of affairs that would not be true of any finite real population. If, however, we work on this assumption, we may consider a graphical representation of the number sick at any particular time as obtained by consideration of two perpendicular axes. Consider axes T_1, epoch of falling sick, and T_2, epoch of incapacity. An individual who falls sick at time T_1 is subsequently sick till time T_2 and would be recorded as contributing one unit to the z ordinate until time T_2 when he would disappear from our consideration. We should thus have, for the number still sick, a hollow monotone surface falling away from the line $T_1 = T_2$. Let us consider this in more detail, and for convenience take the T_1 axis as running eastwards for time increasing and the T_2 one as southwards for time increasing (see Fig. 1). This modification of the usual convention of

axes is for convenience in comparing with our tables, where time increasing is indicated both horizontally to the right and vertically downwards. Then the surface itself will have the following features:

(1) It will be bounded on the north-east by an abrupt precipice at $T_1 = T_2$.

(2) The ridge $T_1 = T_2$ running north-west and south-east will be level (assumption of uniform rate of falling sick).

(3) The surface will tend on the south and west from the ridge asymptotically to zero.

(4) Sections of the surface by any line $T_1 = t$ or $T_2 = t$ will be all congruent monotone J curves, with a maximum where the line intersects the ridge. For we see that at any point P, (t_1, t_2), the ordinate is distant $(t_2 - t_1)$ from the maximum ordinate of the section $T_1 = t_1$ and is distant the same amount $(t_2 - t_1)$ from the equal maximum ordinate of the section $T_2 = t_2$ (see Fig. 1, where $PQ_1 = PQ_2$).

In the school already referred to, the data of Table A were obtained from the experience of seven (non-consecutive) terms free from any pronounced epidemic. This experience covered about 250,000 pupil sessions at risk, being approximately 300 pupils on roll for each of 120 sessions for each of seven terms. There were 1805 spells of absence in the experience, totalling 5915 sessions. Absence from whatever cause was included: nearly always this was due to personal sickness. Every case of absence was noted from its first session to its last. Cases that were absent at the beginning or end of term were not considered, nor, for the sake of simplicity, were those very few of more than 3 weeks' duration: seven such are shown in the 1812 cases dealt with in Table I. There was a little difficulty about holidays in the middle of term. These half-term holidays, etc., usually fell on a Saturday or Monday, so that there is some reduction in the number of cases recorded on these days. Some old cases were, however, ascribed to these days for return if the pupil were absent just before the holiday and were back in attendance immediately afterwards: it was assumed that for such cases the pupil's illness was of the modal duration for his appropriate first session of absence. Cases of pupils leaving the school were observed till their last recorded attendance: they then passed from our experience as no longer at risk. The absences were spread over the week as follows:

TABLE B

Day of week	Mon.		Tues.		Wed.	Thurs.		Fri.		Sat.	Total
Session	M.	A.	M.	A.	M.	M.	A.	M.	A.	M.	
Absences on such sessions	550	578	538	600	589	578	606	596	611	669	5915

TABLE I

Sessions in which absences started and finished

Time of last absence		Time of first absence: Mon. M.	Mon. A.	Tues. M.	Tues. A.	Wed. M.	Thurs. M.	Thurs. A.	Fri. M.	Fri. A.	Sat. M.	Total
Mon.	M.	61	—	—	—	—	—	—	—	—	—	61
	A.	127	71	—	—	—	—	—	—	—	—	198
Tues.	M.	11	6	38	—	—	—	—	—	—	—	55
	A.	20	8	67	81	—	—	—	—	—	—	176
Wed.	M.	43	8	44	17	112	—	—	—	—	—	224
Thurs.	M.	1	1	1	2	6	55	—	—	—	—	66
	A.	7	7	12	6	25	83	66	—	—	—	206
Fri.	M.	2	—	3	2	1	7	3	52	—	—	70
	A.	7	—	3	2	3	20	10	73	58	—	176
Sat.	M.	35	5	19	6	18	40	12	46	19	196	396
Mon.	M.	2	1	3	1	1	—	—	4	1	2	15
	A.	5	—	1	—	4	4	2	7	3	15	41
Tues.	M.	—	—	1	—	—	—	—	—	—	1	2
	A.	2	—	1	—	1	1	1	2	1	3	12
Wed.	M.	3	—	1	—	1	1	—	3	—	7	16
Thurs.	M.	—	—	—	—	—	1	—	1	—	—	2
	A.	—	—	—	—	1	1	—	3	1	1	7
Fri.	M.	—	—	—	—	—	—	—	—	—	—	—
	A.	1	—	—	—	—	1	—	—	—	1	3
Sat.	M.	7	—	6	2	3	11	2	9	1	6	47
Mon.	M.	1	—	—	—	—	1	—	—	—	—	2
	A.	—	—	—	—	—	1	1	—	—	2	4
Tues.	M.	—	—	—	—	—	—	—	—	1	—	1
	A.	1	—	—	—	—	—	—	—	—	—	1
Wed.	M.	1	—	—	—	1	2	—	1	—	—	5
Thurs.	M.	—	—	—	—	—	—	—	—	—	1	1
	A.	—	—	—	—	—	—	—	—	—	1	1
Fri.	M.	—	—	—	—	—	—	—	—	—	—	—
	A.	—	—	—	—	—	—	—	—	—	—	—
Sat.	M.	1	—	3	1	1	5	—	2	—	1	14
Mon.	M.	—	—	—	—	—	—	—	1	—	—	1
	A.	—	—	1	—	—	—	—	—	—	—	1
Tues.	M.	—	—	—	—	—	—	—	—	—	—	—
	A.	1	—	—	—	1	—	—	—	—	—	2
Wed.	M.	—	—	—	—	—	—	—	—	—	—	—
Thurs.	M.	—	—	—	—	—	—	—	—	—	—	—
	A.	—	—	—	—	—	—	—	—	—	—	—
Fri.	M.	—	—	—	—	—	—	—	—	—	—	—
	A.	—	—	—	—	—	—	—	—	1	—	1
Sat.	M.	—	1	2	1	—	—	—	1	—	—	5
Total		339	108	206	121	179	234	97	205	86	237	1812

To a first approximation we find that our assumption that the number absent at any census is constant is justified: all the values approximate to the average of 591·5. There is, however, 13 % excess on Saturday, whilst the afternoons total 2395 against 2262 for the corresponding mornings, a 6 % excess. There is a slight increase during the week, the first four sessions totalling 2266 against 2391 for Thursday and Friday, the increase for the latter again being some 6 %. Monday gives 1128, Tuesday 1138 (1 % increase), Thursday 1184 (4 % increase) and Friday 1207 (2 % increase). We shall, however, at this stage in our enquiry neglect these variations in the rate of absence with epoch, remembering only that we shall not expect any very great precision in our results.

(3) Recording at censuses

In actual practice we do not have machinery for recording each case as it falls sick and as it recovers and ceases to be incapacitated, and the S.R. surface cannot be obtained in the manner that we have just outlined. We have instead records, taken at stated times, of individuals who happen at those epochs to be sick. In a population at large we can call these times censuses: the censuses may be roll calls of military units, register markings of schools, clockings-in of factories, or comparable procedures for other communities. As a result, we have records that can be put in the form of Table I. This states that there were x cases that were not sick at time t_r but were sick at time t_{r+1} who had not recovered at time t_s but had recovered at time t_{s+1}. Thus Table I states that there were sixty-one cases who were first absent on a Monday morning and who made that session their last absence, returning in time for the afternoon register marking. Let us consider in our population, in the general case just mentioned, those who had not recovered at time t_s. These will be the "still sick" of those who fell sick at time t_r (represented by the ordinate at the point $A(t_r, t_s)$ (see Fig. 1)), the still sick of those who fell sick just before time t_{r+1} (represented by the ordinate at the point $B(t_{r+1}-\epsilon, t_s)$), and the still sick of those who fell sick at all intermediate times. In other words, those who have not recovered at time t_s are given by an area on the S.R.C. made by the section of the surface by $T_2 = t_s$, the particular area being that lying between $T_1 = t_r$ and $T_1 = t_{r+1}$. The maximum of this S.R.C. will be at E, where $T_1 = t_s$, and the curve will decrease westwards from the ridge (from E towards BA...). The number that we have just considered will therefore be given by the area of the portion of an S.R.C. for the length of abscissa AB. Similarly, of those who fell sick between t_r and t_{r+1}, the number who had not recovered by time t_{s+1} will be the area of the S.R.C. for the length $\alpha\beta$. We have therefore

$$x = \text{difference in areas of portions of S.R.C. on } AB \text{ and on } \alpha\beta$$
$$= (\text{area on } AE - \text{area on } BE) - (\text{area on } \alpha\epsilon - \text{area on } \beta\epsilon).$$

Again, y, the number of the same batch falling sick between t_r and t_{r+1} who recover between t_{s+1} and t_{s+2}, is given by

$$y = (\text{area on } \alpha\epsilon - \text{area on } \beta\epsilon) - (\text{area on } ae - \text{area on } be).$$

Hence

$$x + y = (\text{area on } AE - \text{area on } BE) - (\text{area on } ae - \text{area on } be).$$

Similarly, if X and Y are the numbers falling sick between t_{r+1} and t_{r+2} who recover between t_s and t_{s+1} and between t_{s+1} and t_{s+2}, respectively, then

$$X + Y = (\text{area on } BE - \text{area on } CE) - (\text{area on } be - \text{area on } ce).$$

Hence

$$x + y + X + Y = (\text{area on } AE - \text{area on } CE) - (\text{area on } ae - \text{area on } ce).$$

We see thus that the cell values are additive, and that for any four censuses the total can be obtained and will give the difference of areas of the S.R.C. corresponding to the various time intervals.

Suppose that we now have any four of our routine censuses t_1, t_2, t_3, and t_4. We can find the number of individuals who, falling sick between t_1 and t_2, recover between t_3 and t_4. Let this be n. Then $n =$ area on standard S.R.C. between $(t_3 - t_1)$ and $(t_3 - t_2)$ less area on same between $(t_4 - t_1)$ and $(t_4 - t_2)$.

Put $t_1 = -\infty$, and $t_4 = +\infty$, so that A, a and c, if the figure is $ACca$, go to infinity. Then

$n =$ area on standard S.R.C. from C at $(t_3 - t_2)$ to the asymptotic end where the ordinate is zero.

We can check this at once by realizing that we have the whole of the curve except that portion lying between C and E: the ordinates all along ac are now zero. We can in this way build up the area of the S.R.C. For we can consider various values of t_3 and of t_2 in turn. We may also note that, as we suggested in dealing with Table A above, the time element enters in in the form of actual lapse of time and not of sessions.

(4) THE CASE OF A DAY SCHOOL

In the case of the school previously referred to the registers are marked in weekly cycles, ten times each week. Let us take as our unit of time the day, and our zero as midnight Sunday/Monday, and let us take epochs in fractions of the day to the nearest first decimal place. We then have observations at the following times:

TABLE C

Day	Mon.		Tues.		Wed.	Thurs.		Fri.		Sat.
Session	M.	A.	M.	A.	M.	M.	A.	M.	A.	M.
Roll call* at Epoch	9.00 0·4	2.20 0·6	9.00 1·4	2.20 1·6	9.00 2·4	9.00 3 4	2.20 3·6	9 00 4·4	2.20 4·6	9.00 5·4

* It may be noted that the *duration* of a session does not here come into consideration. If a pupil fall sick during the session he may be sent home, or made to lie down, but the register is not amended. If he is fit by the next session in such cases there will then be no record. If he is not fit by the next session, he will then be marked absent for the first time. We note further that any case falling sick between two roll calls and recovering in that interval will not be recorded at all. We assume further that the time taken to come to school is negligible—all cases fit at 9 a.m. are recorded as at school and fit, all cases not fit at 9 a.m. are recorded as absent.

We have seen that the absences can be recorded as in Table I. In this we may consider that the entries repeat themselves in weekly cycles along the top diagonal, so that we can read along this ... 61, 71, 38, ..., 58, 196, 61, 71, 38, ..., 58, 196, 61, ... We thus have a table infinitely long along the diagonal, but, as we have already pointed out, giving rise to a sickness recovery surface that will have zero ordinates 21 days away from this diagonal. These zero ordinates will commence therefore along the bottom diagonal, and here the state of affairs is represented in Fig. 2. The surface now lies entirely to the north-east of the diagonal $A_1 A_2 A_3 A_4 A_5$ ... and a_3, for example, could be the one case recorded in Table I as first absent on Tuesday morning (i.e. falling sick in interval epoch $0{\cdot}6d$ to epoch $1{\cdot}4d$) and last absent on the Monday afternoon just short of 3 weeks later (i.e. recovering in interval epoch $21{\cdot}6d$ to $22{\cdot}4d$, where the letter d indicates that the epoch is measured in units of a day). Similarly, b_2 would be the number of cases of the same lot of first absences who recovered between Monday morning roll call (epoch 21·4) and Monday afternoon roll call (epoch 21·6)—in our experience $b_2 = 0$ (see Table I). Then from what we have just seen in the general case c_2, say, is the difference in the areas of sections C_2B_3 and D_2C_3, or what is the same, the difference in the areas of the sections D_2C_2 and C_3B_3. Let us use the notation $|AB|$ to represent the area of the S.R.C on the base AB. Then we have the following relations

$$\left.\begin{array}{l} |A_1B_1| = a_1 \\ |A_2B_2| = a_2 \\ \dots\dots\dots\dots\dots \end{array}\right\}, \qquad \dots\dots(1{\cdot}1)$$

$$\left.\begin{array}{l} |B_1C_1| = b_1 + |A_2B_2| \\ |B_2C_2| = b_2 + |A_3B_3| \\ \dots\dots\dots\dots\dots\dots\dots\dots \end{array}\right\}, \qquad \dots\dots(1\cdot 2)$$

$$\left.\begin{array}{l} |C_1D_1| = c_1 + |B_2C_2| \\ |C_2D_2| = c_2 + |B_3C_3| \\ \dots\dots\dots\dots\dots\dots\dots\dots \end{array}\right\}, \qquad \dots\dots(1\cdot 3)$$

and so on.

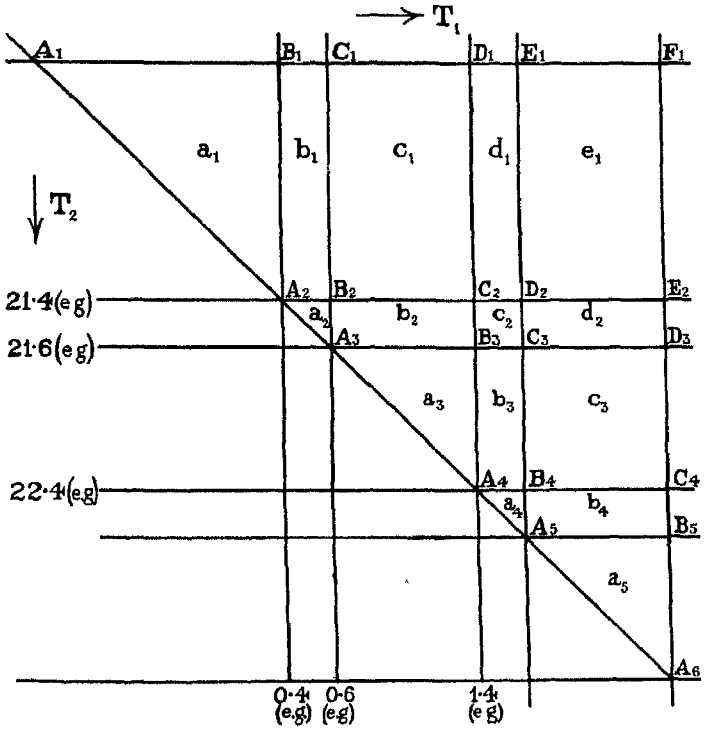


Fig. 2.

Then, adding, we have

$$\left.\begin{array}{l} |A_1B_1| = a_1 \\ |A_1C_1| = a_1 + (a_2 + b_1) \\ |A_1D_1| = a_1 + (a_2 + b_1) + (a_3 + b_2 + c_1) \\ |A_1E_1| = a_1 + (a_2 + b_1) + (a_3 + b_2 + c_1) + (a_4 + b_3 + c_2 + d_1) \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \end{array}\right\}. \qquad \dots\dots(2\cdot 1)$$

In other words, the areas of the S.R.C. on various bases of the section of the S.R. surface by $T_2 = A_1B_1C_1D_1 \dots$, as measured from the westward (zero) end, are given by the accumulated sums of the column totals to this line.

Thus Table I gives us

Last absence on	Recovers between	First absence recorded on				
		Mon. m.	Mon. a.	Tues. m.	Tues. a.	Wed.
		Falls sick between				
		$\bar{2}$·4* and 0·4	0·4 and 0·6	0 6 and 1·4	1·4 and 1·6	1·6 and 2·4
Sat. m.	19·4 and 21·4	1	—	3	1	1
Mon. m.	21·4 and 21 6	—	—	—	—	—
Mon. a.	21·6 and 22·4	—	—	1	—	—
Tues. m.	22·4 and 22·6	—	—	—	—	—
Tues. a.	22·6 and 23·4	—	—	—	—	1

* Read bar 2 plus 0·4, i.e. −1·6 before zero epoch.

If we deal with this in the manner just indicated we have, for the curve along $A_1B_1C_1D_1E_1F_1\ldots$ corresponding to a last absence on the Saturday morning referred to, the series of values

21·0	19·0	18·8	18·0	17·8	17·0	...
0	1	1	5	6	8	...

In the same way we can build up the areas on the sections by various values of T_1, remembering always that the entries of the Table I are the difference between the two parallel walls, in either case, of the rectangular slabs. In the case that we have just considered we have, alternatively, by working along $A_6B_5C_4D_3E_2F_1$ in Fig. 2:

21·0	20·2	20·0	19·2	19·0	17·0	...
0	1	1	2	2	8	...

The result 8 can thus be derived in either of two ways, as is obvious from consideration of the formulae (2·1) and the corresponding ones derived for the T_1 sections.

We can thus prepare Table II. In this we note that the results of the series 0, 1, 1, 5, 6, 8 are entered in the Saturday morning column of the first part, and the second series, 0, 1, 1, 2, 2, 8 are the bottom entries of the column Wednesday morning of the second part. We may note that as already pointed out we do not have any readings for triangular portions such as Q_1PQ_2 of Fig. 1, but these are not needed for the computation of the areas of the various sections of the S.R. surface.

The method used for obtaining Table II is of wide generality and can be used at whatever irregular intervals we may take our censuses.

In our case, as we have the decade of roll calls per week, the pattern of entries repeats every 7 days down any column. As, further, our roll calls are all at 9 a.m. (mornings) and 2.20 p.m. (afternoons) the intervals recur at intervals of $0{\cdot}2d$, $0{\cdot}8d$, $1{\cdot}0d$,...etc., and in consequence we have a number of entries (possibly 20 and at least 6), for each interval from the beginning. These entries, for any one interval, should all be equal, but of course in fact, owing to random variations, they differ somewhat. They do however run very closely, in the main, to the same order of magnitude. We note, however, that the T_1 entries increase steadily to the right, owing to the slight increase already reported of "incapacity" during the week, and in particular to the 13 % excess on Saturday morning. We shall take the average values of the columns to give the best value for the area of the S.R.C. of the population considered from any abscissa to the tail. Actually we note that these averages show certain irregularities—thus at $14{\cdot}0d$ and at $17{\cdot}0d$ there are increases in the area, whilst on taking differences that from $1{\cdot}2$ to $1{\cdot}8$ is less, over a larger range, than that from $1{\cdot}0$ to $1{\cdot}2$ or from $1{\cdot}8$ to $2{\cdot}0$. We shall not trouble to graduate the result, but proceed immediately to use it to reform Table I, working backwards through the process used to build up Table II. The more important features are given below:

TABLE D

		Mon.		Tues.		Wed.	Thurs.		Fri.		Sat.	Total
		M.	A.	M.	A.	M.	M.	A.	M.	A.	M.	
Cases absent for one session only	Expected	69	55	55	55	139	63	55	55	55	139	740
	Actual	61	71	38	81	112	55	66	52	58	196	790
	$d_1 = A - E$.	-8	$+16$	-17	$+26$	-27	-8	$+11$	-3	$+3$	$+57$	$+50$
Total cases of absence starting on session stated	Expected	332	95	200	95	200	239	95	200	95	200	1751
	Actual	338	107	204	120	179	234	97	204	85	237	1805
	$d_2 = A - E$.	$+6$	$+12$	$+4$	$+25$	-21	-5	$+2$	$+4$	-10	$+37$	$+54$
	$d_1 - d_2$	$+14$	$+4$	-21	$+1$	-6	-3	$+9$	-7	$+13$	$+20$	

Again we observe the great excess of Saturday morning cases, many for one session only. It may be noted that this is in spite of special precautions taken at the school in question where this tendency was originally a very pronounced tradition: one method that was found successful with the keener pupils was to

TABLE II

Area of tail of sickness recovery curve from abscissae of specified value. Twenty different estimates

Time interval (days)	Abscissa parallel to T_1 axis										Abscissa parallel to T_2 axis										Av.
	Mon.		Tues.		Wed.	Thurs.		Fri.		Sat.	Mon.		Tues.		Wed.	Thurs.		Fri.		Sat.	
	M.	A.	M.	A.	M.	M.	A.	M.	A.	M.	M.	A.	M.	A.	M.	M.	A.	M.	A.	M.	
0·0	550	578	538	600	589	578	606	596	611	669	550	578	538	600	589	578	606	596	611	669	591·5
0·2	—	471	—	480	—	—	509	—	526	—	471	—	480	—	—	509	—	526	—	—	496·5
0·8	—	—	334	—	410	—	—	392	—	432	—	334	—	410	—	—	392	—	432	—	392·0
1·0	—	—	298	314	371	344	—	361	374	405	298	314	371	—	344	361	374	405	—	—	352·4
1·2	—	—	—	284	—	—	330	—	346	—	284	—	—	—	330	346	—	—	—	—	320·0
1·8	—	—	—	—	272	277	—	—	—	326	—	272	—	277	—	—	326	—	—	—	291·6
2·0	212	—	—	—	250	255	269	265	—	308	250	—	255	269	265	308	—	—	—	212	259·8
2·2	—	194	—	—	—	—	249	—	257	—	—	—	249	—	257	—	—	—	—	194	233·3
2·8	171	—	—	—	—	200	—	229	—	—	—	200	—	229	—	—	—	—	171	—	200·0
3·0	163	155	148	—	—	186	195	215	222	239	186	195	215	222	239	—	—	163	155	148	190·4
3·2	—	148	—	145	—	—	182	—	210	—	182	—	210	—	—	—	—	148	—	145	171·2
3·8	130	—	124	—	—	—	—	173	—	207	—	173	—	207	—	—	130	—	124	—	158·5
4·0	124	119	120	122	131	—	—	167	171	197	167	171	197	—	—	124	119	120	122	131	143·9
4·2	—	113	—	118	—	—	—	—	165	—	165	—	—	—	—	113	—	118	—	—	132·0
4·8	—	—	98	—	111	—	—	—	—	161	—	161	—	—	—	—	98	—	111	—	123·3
5·0	95	—	94	96	108	110	—	—	—	155	155	—	—	—	95	94	96	108	—	110	109·7
5·2	—	84	—	92	—	—	107	—	—	—	—	—	—	—	84	92	—	—	—	107	94·3
5·8	81	—	—	—	88	97	—	—	—	—	—	—	—	81	—	—	88	—	97	—	88·7
6·0	77	71	69	—	85	94	94	99	—	—	—	—	77	71	69	85	—	94	94	99	84·1
6·2	—	68	—	67	—	—	91	—	99	—	—	—	68	—	67	—	—	91	—	99	81·2
6·8	60	—	60	—	—	77	—	87	—	—	—	60	—	60	—	—	77	—	87	—	71·0
7·0	59	54	57	58	61	74	75	85	87	96	59	54	57	58	61	74	75	85	87	96	70·6
7·2	—	54	—	55	—	—	72	—	85	—	54	—	55	—	—	72	—	85	—	—	66·5
7·8	—	—	44	—	53	—	—	72	—	85	—	44	—	53	—	—	72	—	85	—	63·5
8·0	—	—	44	43	50	51	—	69	72	83	44	43	50	—	51	69	72	83	—	—	58·9
8·2	—	—	—	43	—	—	50	—	69	—	43	—	—	—	50	69	—	—	—	—	54·0
8·8	—	—	—	—	39	44	—	—	—	70	—	39	—	44	—	—	70	—	—	—	51·0

TABLE II (*continued*)

9·0	35	—	—	—	39	41	43	48	—	67	39	—	41	43	48	67	—	—	—	35	45·5
9·2	—	32	—	—	—	—	40	—	48	—	—	—	40	—	48	—	—	—	—	32	40·0
9·8	30	—	—	—	—	31	—	42	—	—	—	31	—	42	—	—	—	—	30	—	34·3
10·0	29	27	27	—	—	31	30	39	42	47	31	30	39	42	47	—	—	29	27	27	34·0
10·2	—	26	—	26	—	—	30	—	39	—	30	—	39	—	—	—	—	26	—	26	30·2
10·8	25	—	24	—	—	—	—	29	—	41	—	29	—	41	—	—	25	—	24	—	29·7
11·0	24	22	23	23	24	—	—	29	29	38	29	29	38	—	—	24	22	23	23	24	26·5
11·2	—	21	—	23	—	—	—	—	29	—	29	—	—	—	—	21	—	23	—	—	24·3
11·8	—	—	19	—	21	—	—	—	—	28	—	28	—	—	—	—	19	—	21	—	22·7
12·0	15	—	19	19	21	19	—	—	—	28	28	—	—	—	15	19	19	21	—	19	20·0
12·2	—	13	—	19	—	—	18	—	—	—	—	—	—	—	13	19	—	—	—	18	16·7
12·8	12	—	—	—	17	16	—	—	—	—	—	—	—	12	—	—	17	—	16	—	15·0
13·0	11	10	12	—	17	16	16	17	—	—	—	—	11	10	12	17	—	16	16	17	14·1
13·2	—	9	—	12	—	—	16	—	17	—	—	—	9	—	12	—	—	16	—	17	13·5
13·8	7	—	9	—	—	13	—	16	—	—	—	7	—	9	—	—	13	—	16	—	11·2
14·0	7	5	8	9	10	13	13	16	16	17	7	5	8	9	10	13	13	16	16	17	11·4
14·2	—	5	—	8	—	—	13	—	16	—	5	—	8	—	—	13	—	16	—	—	10·5
14·8	—	—	4	—	7	—	—	13	—	16	—	4	—	7	—	—	13	—	16	—	10·0
15·0	—	—	4	4	6	8	—	13	13	16	4	4	6	—	8	13	13	16	—	—	9·1
15·2	—	—	—	4	—	—	8	—	13	—	4	—	—	—	8	13	—	—	—	—	8 3
15·8	—	—	—	—	2	6	—	—	—	13	—	2	—	6	—	—	13	—	—	—	7·0
16·0	3	—	—	—	2	5	6	8	—	13	2	—	5	6	8	13	—	—	—	3	6·2
16·2	—	2	—	—	—	—	5	—	8	—	—	1	5	—	8	—	—	—	—	2	5·0
16·8	3	—	—	—	—	1	—	6	—	—	—	1	—	6	—	—	—	—	3	—	3·3
17 0	3	2	1	—	—	1	1	5	6	8	1	—	5	6	8	—	—	3	2	1	3·4
17·2	—	2	—	1	—	—	1	—	5	—	1	1	5	—	—	—	—	2	—	1	2 2
17·8	2	—	1	—	—	—	—	1	—	6	—	1	—	6	—	—	2	—	1	—	2·5
18·0	2	2	1	1	0	—	—	1	1	5	1	—	—	—	—	2	2	1	1	0	1·6
18·2	—	2	—	1	—	—	—	—	1	—	1	1	—	—	—	2	—	1	—	—	1·3
18·8	—	—	1	—	0	—	—	—	—	1	—	1	—	—	—	—	1	—	0	—	0·7
19·0	2	—	1	1	0	0	—	—	—	1	1	—	—	—	2	1	1	0	—	0	0·8
19 2	—	2	—	1	—	—	0	—	—	—	—	—	—	—	2	1	—	—	—	0	1·0
19·8	1	—	—	—	0	0	—	—	—	—	—	—	—	1	—	—	0	—	0	—	0 3
20 0	1	1	1	—	0	0	0	0	—	—	—	—	1	1	1	0	—	0	0	0	0·4
20·2	—	1	—	1	—	—	0	—	0	—	—	—	1	—	1	—	—	0	—	0	0·5
20·8	0	—	0	—	—	0	—	0	—	—	—	0	—	0	—	—	0	—	0	—	0·0
21·0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0·0

hold the headmaster's terminal examinations week by week on Saturdays. We note further the tendency, again more pronounced for the one session cases, for absence in the afternoons. I fancy that the reason is that if Tommy feels a slight malaise in the afternoon his mother (it is the mother who settles these matters) says, "Well, it's only three lessons this afternoon and one of them is P.T. and one is Art. He needn't go." The same malaise would not be enough to keep him at home for the longer and harder morning session. It would be of interest to note if this feature of afternoon absences were more in evidence in schools with a very short afternoon session than in the one of our experience (mornings 3 hr. 20 min., afternoons 2 hr. 10 min.).

On the other hand, we note that there are a number who attend on Wednesdays who apparently would not come that session if it began a full day: the attitude may be: "Well, I'll see it out this morning, and can go to bed this afternoon if I am not better then." This introduces a practical consideration in organization, for in general the more nearly equal we make our census intervals the greater is the total absence recorded (this follows as an easy corollary from the assumption of a monotone hollow decreasing S.R.C.) so that a 6 morning-4 afternoon week would in the ordinary way give more absences than a 5 morning-5 afternoon week. If however Wednesday were a full day, then perhaps some of those who now attend on Wednesday would fail to do so. There may be of course some countervailing effect on the attendance for the last session of the week. I am not aware of any data comparable to ours for a 5-day week school. It is relevant to note here that administrators have expressed some concern at the psychological effects (in excessive "week-ending", etc.) of the 5-day week (see, e.g., J. S. Hart, *J.S.S.* vol. LXXXV, pt III, pp. 349–411, May 1922). We should point out that in none of the terms of our experience were there more than two Jews (observing Saturday sabbath) in the school and for most of the time there was none.

(5) Sickness recovery as experienced

By differencing the areas of the last column of Table II we could obtain the ordinates of the S.R.C. As, however, the areas are not smoothed the errors will be, as already pointed out, noticeable at an early stage, and no great reliance can be put on the results. The differences for important bands of width $0{\cdot}2d$ are given by the following summary:

TABLE E

Beginning at... d	0·0	1·0	2·0	3·0	4·0	7·0	14·0	21·0
Difference	95·0	32·4	26·5	19·2	11·9	4·1	0·9	0·0
$l_x = 100$	100	34	28	20	12·5	4	1	0

We can conclude that recovery for these adolescent boys and girls is very rapid: two-thirds of them are fit again in 24 hr., three-quarters in 2 days, and only 1 % will be more than a fortnight absent.

(6) SUMMARY AND CONCLUSIONS

1. From observations at irregular intervals of the numbers of a population absent, a method is given for obtaining the curve of recovery from sickness (or whatever the cause of absence may be).
2. The sickness recovery curve is a monotone curve, and the total of absences is least for a definite number of censuses in any time if the censuses are equally spaced.
3. The observations of numbers absent and recovering at various censuses give portions of the area of the S.R.C.: we do not know and need not have the number who both fall sick and recover in any intercensal interval.
4. The method is applied to a secondary school population of about a quarter of a million pupil sessions at risk, and the actual and predicted absences compared.
5. Some administrative aspects of the intervals between register markings are pointed out.
6. For the particular adolescent boys and girls, the conclusion is reached that two-thirds of them are fit within one day, and that generally recovery is very rapid.

THE SIGNIFICANCE OF THE DIFFERENCE BETWEEN TWO MEANS WHEN THE POPULATION VARIANCES ARE UNEQUAL

By B. L. WELCH, Ph.D.

1. *Introduction.* Suppose that we have samples of sizes n_1 and n_2 from populations π_1 and π_2 respectively. Let the populations be normal in form, π_1 having mean and standard deviation α_1 and σ_1, and π_2 having mean and standard deviation α_2 and σ_2. Let it be required to test whether $\alpha_1 = \alpha_2$. Two cases may be distinguished: (i) σ_1 and σ_2 may be equal or (ii) they may be unequal. In the first case the most appropriate test for the equality of the α's is made by referring the criterion

$$u = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\Sigma_1 + \Sigma_2}{(n_1 + n_2 - 2)}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad \ldots\ldots(1)$$

to the t distribution with $f = (n_1 + n_2 - 2)$.* In the second case, if the ratio of the two σ's is known, a similar criterion can be used: if, however, this ratio is unknown, no criterion quite so simple is available. A solution of the problem of testing the hypothesis in this instance has been proposed by R. A. Fisher,† using the concept of fiducial distributions. Fisher notes the equivalence of his test to that given previously by W. V. Behrens‡ in 1929. The validity of this test has, however, been questioned by M. S. Bartlett.§ An alternative|| criterion which has been often employed is

$$v = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\Sigma_1}{n_1(n_1 - 1)} + \frac{\Sigma_2}{n_2(n_2 - 1)}}}. \qquad \ldots\ldots(2)$$

This may be referred to the normal probability table if the samples are large enough, but for small samples it does not yield an exact test and it is not clear how it may best be made to furnish approximations.

It has been pointed out by Fisher that in many practical situations where u is used, the fact that the σ's must be equal for the criterion to be distributed as t does not necessarily mean that an *assumption* of equality is involved. It may

* Σ_1 denotes the sum of squares of the observations in the first sample from their mean. Σ_2 similarly for the second sample.

† *Ann. Eugen.* VI, Part IV (1936), p. 396.

‡ *Landw. Jb.* LXVIII (1929), p. 822.

§ M. S. Bartlett. *Proc. Camb. Phil. Soc.* XXXII, Part 4 (1936), p. 564.

|| It should be noted that, if $n_1 = n_2$, the criteria u and v are identical.

mean that the equality of the σ's is being regarded as part of the hypothesis under test. In such situations it may be argued that there is no point in testing whether $\alpha_1 = \alpha_2$ unless we have also $\sigma_1 = \sigma_2$. However, even if the question posed is one of testing whether two normal populations are identical, u will not necessarily be the best criterion to use. u will afford a valid* test, in the sense that it will control satisfactorily the chance of rejecting the hypothesis when it is actually true, but it is only one of many such. The choice of criterion must depend on what sort of departure from the hypothesis under test we are most interested in detecting. u is demonstrably the best criterion when we wish to detect differences in means without attendant differences in standard deviations. It is conceivable, however, that the test based on u may sometimes operate in such a fashion that differences in the standard deviations σ_1 and σ_2 may mask differences in the means α_1 and α_2, with the result that judgments of non-significance may be too frequently made. The investigations in this paper throw some light on this point, although explicitly they are concerned with cases where it is reasonable to test whether $\alpha_1 = \alpha_2$, whatever the ratio of σ_1 to σ_2.

In the first place I shall consider the problem—how far is the criterion u valid even when $\sigma_1 \neq \sigma_2$? (That the test is liable to be biased in this instance is generally realized, but the extent of the bias has not hitherto received any detailed discussion.) In the second place I shall consider the validity of testing the hypothesis by referring v to the t distribution with $f = (n_1 + n_2 - 2)$. Finally, I wish to make some observations about the test of Fisher and Behrens, mentioned above.

It is easily seen that u in general is not distributed as t. For whereas the square of the standard error of $(\bar{x}_1 - \bar{x}_2)$ is $(\sigma_1^2/n_1 + \sigma_2^2/n_2)$, the quantity under the root in (1) is an unbiased estimate of

$$\frac{(n_1 - 1)\,\sigma_1^2 + (n_2 - 1)\,\sigma_2^2}{(n_1 + n_2 - 2)}\left(\frac{1}{n_1} + \frac{1}{n_2}\right).$$

This is equal to $(\sigma_1^2/n_1 + \sigma_2^2/n_2)$ only if $\sigma_1 = \sigma_2$ or $n_1 = n_2$. The criterion v does not suffer from this objection, but its distribution still depends to a certain extent on σ_1/σ_2. The first problem will be to obtain the distributions of u and v. The exact distributions will not be derived here, but only certain approximations adequate, I believe, for the purpose in hand.

2. *The distributions of u and v.* When $\alpha_1 = \alpha_2$ we may write

$$(\bar{x}_1 - \bar{x}_2) = \chi' \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}; \quad \Sigma_1 = \chi_1^2 \sigma_1^2; \quad \Sigma_2 = \chi_2^2 \sigma_2^2, \qquad \ldots\ldots(3)$$

where χ'^2, χ_1^2 and χ_2^2 are independently distributed as χ^2 with degrees of freedom 1,

¶ The term "validity" applied to a test is used throughout this paper in the sense here indicated The term "unbiased" is also used with the same meaning, which should not be confused with the meaning which J. Neyman and E. S Pearson have attached to it in recent papers on testing statistical hypotheses.

(n_1-1) and (n_2-1) respectively. It is therefore possible to write both u and v in the form

$$y=\frac{\chi'}{\sqrt{a\chi_1^2+b\chi_2^2}}=\frac{\chi'}{\sqrt{w}}\ \text{(say)}, \qquad\qquad \text{......(4)}$$

where a and b are constants depending on the n's and σ's. w is always distributed independently of χ' and, when $a=b$ or when either a or b is zero, w is distributed as χ^2 multiplied by some constant. In these cases y will be distributed as t multiplied by some constant. For other values of a and b the distribution is not so simple, but a useful approximation may be obtained. Following the lines adopted in a previous paper* let us first approximate to the distribution of w by the Pearsonian Type III Curve

$$p(w)=\frac{1}{(2g)^{\frac{1}{2}f}\,\Gamma(\frac{1}{2}f)}\,w^{\frac{1}{2}f-1}\,e^{-\frac{w}{2g}}, \qquad\qquad \text{......(5)}$$

where f and g are so chosen that the first two moments of the curve agree with the true moments of w. For the curve we have

$$\text{mean}=gf; \quad \mu_2=2g^2f$$

and for the true moments of w

$$\text{mean}=(af_1+bf_2); \quad \mu_2=2\,(a^2f_1+b^2f_2),$$

f_1 and f_2 now being written instead of (n_1-1) and (n_2-1). Hence

$$g=\frac{a^2f_1+b^2f_2}{af_1+bf_2}; \quad f=\frac{(af_1+bf_2)^2}{a^2f_1+b^2f_2}. \qquad\qquad \text{......(6)}$$

With these values of f and g we see from (5) that w/g is distributed approximately as χ^2 with f degrees of freedom. Hence χ' divided by $\sqrt{w/fg}$ is distributed approximately as t. Therefore from (4) we have $y=ct_f$, where

$$c=\frac{1}{\sqrt{fg}}=\frac{1}{\sqrt{af_1+bf_2}}, \qquad\qquad \text{......(7)}$$

and t_f is distributed approximately as t with degrees of freedom f† given by (6). This approximation is sufficiently close for the purpose of the comparisons made in this paper‡ and it will be used throughout. The term "approximation" will be omitted.

From (1) and (3) it is seen that u is of the form (4), where

$$a=\frac{\sigma_1^2\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}{(n_1+n_2-2)\left(\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}\right)}; \quad b=\frac{\sigma_2^2\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}{(n_1+n_2-2)\left(\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}\right)}.$$

* B. L. Welch. *J. Roy. Statist. Soc.* Supplement III, No. 1 (1936). p. 47.

† f is, of course, not now necessarily an integral number of degrees of freedom. It is simply a term in a mathematical approximation. This approximation is of the same form as a true t distribution and hence we may regard f as effectively a number of degrees of freedom.

‡ For some further discussion of the adequacy of this approximation, see a Note by Miss E. Tanburn at the end of this paper.

Hence from (6) and (7) we have $u=ct_f$, where

$$f=\frac{\{(n_1-1)\sigma_1^2+(n_2-1)\sigma_2^2\}^2}{\{(n_1-1)\sigma_1^4+(n_2-1)\sigma_2^4\}};\quad c=\sqrt{\frac{(n_1+n_2-2)\left(\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}\right)}{\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)\{(n_1-1)\sigma_1^2+(n_2-1)\sigma_2^2\}}}. \quad \ldots\ldots(8)$$

Similarly for v we have

$$a=\frac{\dfrac{\sigma_1^2}{n_1(n_1-1)}}{\left(\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}\right)};\quad b=\frac{\dfrac{\sigma_2^2}{n_2(n_2-1)}}{\left(\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}\right)},$$

and we can write $v=ct_f$, where

$$f=\frac{\left(\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}\right)^2}{\dfrac{\sigma_1^4}{n_1^2(n_1-1)}+\dfrac{\sigma_2^4}{n_2^2(n_2-1)}};\quad c=1. \quad \ldots\ldots(9)$$

3. *The validity of the criterion u.* Suppose that u is being used to test the hypothesis that $\alpha_1=\alpha_2$ and that the risk of rejecting the hypothesis when true is to be fixed at some prescribed level ϵ. If it can be assumed that $\sigma_1=\sigma_2$, then from t tables with (n_1+n_2-2) degrees of freedom it is possible to choose u_0 such that the chance $P(|u|>u_0)=\epsilon$. If u_0 is so chosen, but it happens that $\sigma_1\neq\sigma_2$, then the test, which consists in rejecting the hypothesis when $|u|>u_0$, will be biased. We shall have

$$P(|u|>u_0)=P(|ct_f|>u_0)=P\left(|t_f|>\frac{u_0}{c}\right), \quad \ldots\ldots(10)$$

where c and f are given by (8). Owing to the connection between the t distribution and the Beta-function it can be shown that

$$P(|t_f|>t_0)=I_{\frac{f}{f+t_0^2}}(\tfrac{1}{2}f,\tfrac{1}{2}),$$

where
$$I_{z_0}(p,q)=\frac{1}{B(p,q)}\int_0^{z_0}z^{p-1}(1-z)^{q-1}dz.$$

Hence from (10)
$$P(|u|>u_0)=I_{z_0}(\tfrac{1}{2}f,\tfrac{1}{2}), \quad \ldots\ldots(11)$$

where
$$z_0=\frac{f}{\left(f+\dfrac{u_0^2}{c^2}\right)}. \quad \ldots\ldots(12)$$

Since for given sample sizes, c and f depend only on the ratio $\theta=\sigma_1^2/\sigma_2^2$, it is possible from (11) and (12) to obtain for any θ the chance of rejection of the hypothesis when it is true. This dependence on θ is best illustrated by taking particular examples.

Example I. Let $n_1 = n_2 = 10$. Suppose the chance of rejection ϵ is to be fixed at 0·05. The value of u_0 appropriate when $\theta = 1$ is found to be 2·101. In this case, as always when the n's are equal, c is unity. f is $9(\theta+1)^2/(\theta^2+1)$. The values of $P(|u| > 2{\cdot}101)$ for different θ were obtained from (11), using the Incomplete Beta-function Tables,* and are plotted in Fig. 1 as curve (a). For convenience θ is on a logarithmic scale. It is seen that P always lies between 0·05 and 0·065, the latter value being attained when the variation in one of the populations is zero The test is therefore never very seriously biased.

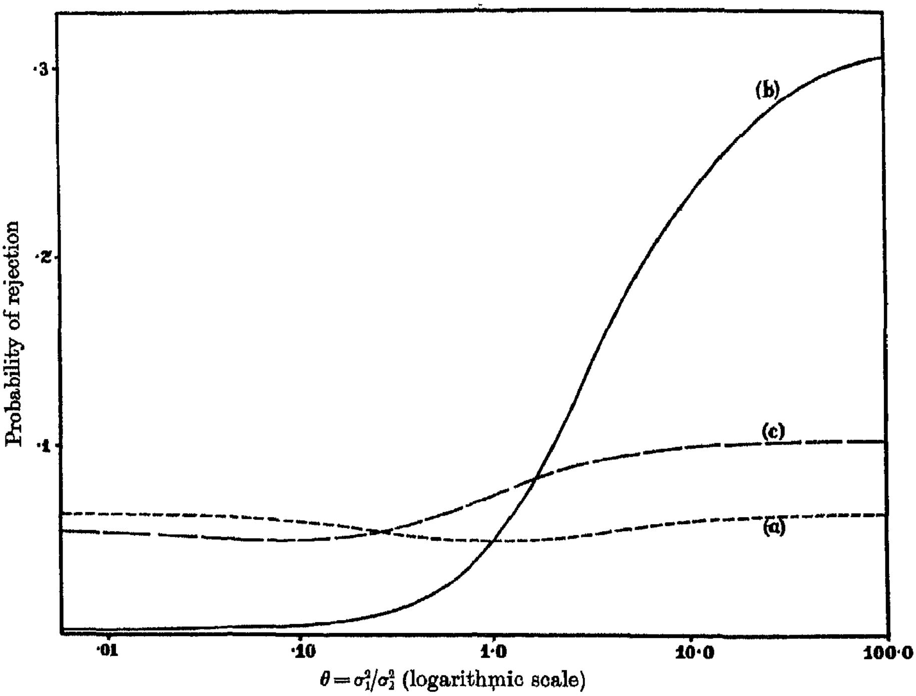


Fig. 1. Probability of rejection of hypothesis $\alpha_1 = \alpha_2$ when true plotted against θ. (a) $n_1 = n_2 = 10$. $P(|u| > 2{\cdot}101)$; (b) $n_1 = 5$, $n_2 = 15$, $P(|u| > 2{\cdot}101)$; (c) $n_1 = 5$, $n_2 = 15$, $P(|v| > 2{\cdot}101)$.

Example II. Let $n_1 = 5, n_2 = 15, \epsilon = 0{\cdot}05$. In this case $(n_1 + n_2 - 2)$ is 18 as before and $u_0 = 2{\cdot}101$. (8) gives

$$f = \frac{(4\theta + 14)^2}{(4\theta^2 + 14)}; \quad c^2 = \frac{18(3\theta + 1)}{4(4\theta + 14)}.$$

$P(|u| > 2{\cdot}101)$ is plotted against θ in Fig. 1 (b). It will now be seen that P varies from 0·0024 when $\theta = 0$ to 0·05 when $\theta = 1$ and then to 0·313 when $\theta = \infty$. There is therefore the possibility of a considerable bias in the test. The significance of the difference between the two means will tend to be under-estimated when $\sigma_1 < \sigma_2$

* *Tables of the Incomplete Beta-function*, edited by Karl Pearson, Biometrika Office, University College, London (1934).

and overestimated when $\sigma_1 > \sigma_2$. The reason for this is not so much that f may differ from 18, but that c can differ considerably from unity. In general the greater the disparity between the n's the more likely is this c factor to bias the test. For equal-sized samples, except perhaps when they are as small as two, the test is never very much biased, whatever θ.

4. *The validity of the criterion* v. The validity of the procedure of testing the hypothesis by referring the criterion v to the t distribution with (n_1+n_2-2) degrees of freedom may be investigated in the same manner. When the n's are equal there is no need for separate discussion as u and v are then identical. Let us consider the case $n_1=5$, $n_2=15$. We find

$$f=\frac{28\,(3\theta+1)^2}{(63\theta^2+2)}; \quad c=1.$$

$P(|v|>2{\cdot}101)$, obtained from an equation similar to (11), is plotted against θ in Fig. 1 (c). As θ increases from 0 to 2/21 P decreases from 0·054 to 0·050 and then increases again to 0·104 at $\theta=\infty$. It is seen that the test formulated in this way is only unbiased when $\theta=2/21$. There is not, however, the possibility of so large a bias as occurs for some values of θ when using the u criterion. The reason of course is that c is always unity, this being guaranteed by the fact that the expectation of the square of the denominator of v is $(\sigma_1^2/n_1+\sigma_2^2/n_2)$. Bias is due solely to f being in general less than 18. When $\theta=0$, f is only 14. f increases to 18 at $\theta=2/21$ and then decreases to 4 at $\theta=\infty$. When the smaller sample comes from the more variable population the effective number of degrees of freedom f is liable to be much smaller than (n_1+n_2-2). Even when $\theta=1$ the effective degrees of freedom in the present case are 6·89, as against 18 for u. If it is known that $\sigma_1=\sigma_2$, then there can be no doubt that u is a better, more sensitive,* criterion than v. If, however, there exists the possibility that σ_1 and σ_2 differ, then u may give very misleading results and it will be safer to use v.†

5. *The comparison of regressions*. It has been found that a criterion based on an estimate of an assumed common variance may lead to a biased test if n_1 and n_2 are different. In the majority of cases it is possible to arrange that n_1 and n_2 are equal or almost so, and hence, practically, serious errors of the kind discussed in the previous sections will not often occur. But there is a more general class of problem where the assumption of equal variances may lead to trouble. An instance is afforded by the test for the equality of linear regression coefficients. Here the usual criterion is

$$u=\frac{(b_1-b_2)}{\sqrt{\dfrac{(\Sigma_1+\Sigma_2)}{(n_1+n_2-4)}\left(\dfrac{1}{\Sigma(x-\bar{x}_1)^2}+\dfrac{1}{\Sigma(x-\bar{x}_2)^2}\right)}},$$

* For further discussion of what is meant here by "sensitivity", see § 6.

† Fig. 1 clearly shows that θ need not differ much from unity before v becomes less biased than u. (This refers of course to the particular sample sizes $n_1=5$, $n_2=15$.)

where b_1 and b_2 are sample regressions and Σ_1 and Σ_2 are now sums of squares from the fitted regression straight lines. Considerations similar to the above show that even if the sample sizes are equal, unless $\Sigma(x-\bar{x}_1)^2$ and $\Sigma(x-\bar{x}_2)^2$ are also equal, the test is biased when the residual variances about the two population regression lines are not the same.

More generally suppose that we have a situation where the samples yield independent statistics T_1, Σ_1, T_2, Σ_2. Let T_1 be normally distributed about α_1 with standard deviation $\sqrt{V_1}\sigma_1$ and Σ_1 be distributed as $\chi_1^2\sigma_1^2$ with f_1 degrees of freedom. Let T_2 be normally distributed about α_2 with standard deviation $\sqrt{V_2}\sigma_2$ and Σ_2 be distributed as $\chi_2^2\sigma_2^2$ with f_2 degrees of freedom. To test whether $\alpha_1=\alpha_2$ we may use the criterion

$$u=\frac{(T_1-T_2)}{\sqrt{\dfrac{(\Sigma_1+\Sigma_2)(V_1+V_2)}{(f_1+f_2)}}}.$$

This is appropriate if $\theta=1$. Otherwise u is approximately distributed as ct_f, where

$$f=\frac{(f_1\theta+f_2)^2}{(f_1\theta^2+f_2)};\quad c^2=\frac{(f_1+f_2)(V_1\theta+V_2)}{(f_1\theta+f_2)(V_1+V_2)}.$$

The effective number of degrees of freedom is (f_1+f_2) only when $\theta=1$. The chief cause of bias, however, is likely to arise from the c^2 factor. When $\theta=0$, c^2 is $V_2(f_1+f_2)/f_2(V_1+V_2)$ and when $\theta=\infty$, c^2 is $V_1(f_1+f_2)/f_1\cdot(V_1+V_2)$. c^2 increases or decreases steadily between these limits according as V_1/f_1 is greater or less than V_2/f_2. c^2 is only uniformly equal to unity if V_1 and V_2 are inversely proportional to f_1 and f_2. In the simple regression case f_1 and f_2 are (n_1-2) and (n_2-2). V_1 and V_2 are $1/\Sigma(x-\bar{x}_1)^2$ and $1/\Sigma(x-\bar{x}_2)^2$. When $n_1=n_2$, c^2 is uniformly unity for all θ only if $\Sigma(x-\bar{x}_1)^2=\Sigma(x-\bar{x}_2)^2$.

The alternative criterion which makes use of separate estimates of σ_1^2 and σ_2^2 is

$$v=\frac{(T_1-T_2)}{\sqrt{\dfrac{V_1\Sigma_1}{f_1}+\dfrac{V_2\Sigma_2}{f_2}}}.$$

This leads to

$$f=\frac{(V_1\theta+V_2)^2}{\left(\dfrac{V_1^2\theta^2}{f_1}+\dfrac{V_2^2}{f_2}\right)};\quad c=1. \qquad \ldots\ldots(13)$$

Any bias is now due only to the effective number of degrees of freedom which is never less than the smaller of f_1 and f_2. It is clear that in certain situations where a criterion of the u type is customarily used, the condition $\theta=1$ needs to be satisfied very stringently. A criterion of the v type will be much safer.

6. *Choice of effective number of degrees of freedom for v.* The question remains—if v is used, what is the best value to take for f? In the above discussion the

consequences of referring v to t tables with (f_1+f_2) degrees have been considered. It was seen that this was absolutely valid only if θ had a particular value. (In the example of Fig. 1 this was $\theta=2/21$. In general it is $\theta=V_2f_1/V_1f_2$.) When there is strong *a priori* reason for believing that θ is in the neighbourhood of a certain value, then it will be better to take the f obtained by substituting this value in (13). For instance suppose that we have reason to believe that $\theta \simeq 1$. Then, for the v test, it will be preferable to take

$$f=\frac{(V_1+V_2)^2}{\left(\frac{V_1^2}{f_1}+\frac{V_2^2}{f_2}\right)}. \qquad \ldots\ldots(14)$$

In the example of section 4 this gives $f=6{\cdot}89$ and the corresponding critical value v_0 is $2{\cdot}374$. In Fig. 2 a comparison is made for different θ between the two rules:

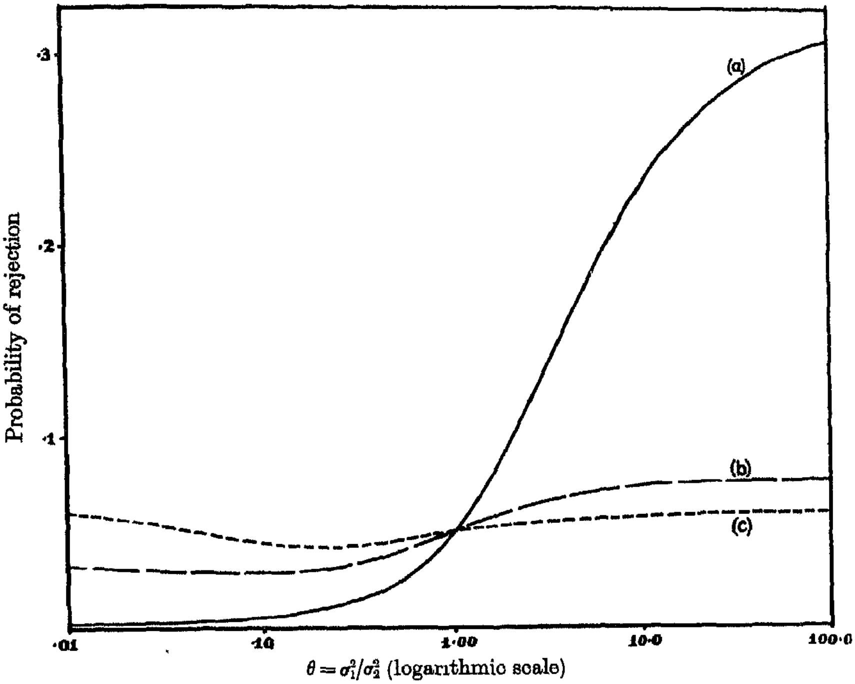


Fig. 2. Probability of rejection of hypothesis $\alpha_1=\alpha_2$ when true plotted against θ. (*a*) $n_1=5$, $n_2=15$, $P(|u|>2{\cdot}101)$; (*b*) $n_1=5$, $n_2=15$, $P(|v|>2{\cdot}374)$; (*c*) $n_1=5$, $n_2=15$, $P(|z|>1{\cdot}861)$.

(*a*) reject the hypothesis that $\alpha_1=\alpha_2$ if $|u|>2{\cdot}101$ and (*b*) reject if $|v|>2{\cdot}374$. As arranged, now, both these rules have the property that for $\theta=1$, the chance of rejection of the hypothesis when true is 0·05.

It may be objected that it is illogical to use the v criterion and at the same time regard it as having effective degrees of freedom f given by substituting $\theta=1$ in (13). For, if $\theta=1$, it is known that the u test is better from the point of view of sensitivity, i.e. any real difference $(\alpha_1-\alpha_2)$ will then be detected more frequently

by u than by v. However in taking the value of f given by (14) we are not assuming that θ is *exactly* unity, but simply making use of our reasons for believing that it is near unity. The v test based on this value of f is biased when $\theta \neq 1$ but the bias is seen to be less than that of the u test. The small* advantage which u enjoys with respect to sensitivity when $\theta = 1$ is soon offset by this gain of v in controlling the chance of rejecting the hypothesis that $\alpha_1 = \alpha_2$ when it is actually true.

When there is no very precise *a priori* information about θ available it might seem permissible to use the ratio Σ_1/Σ_2 from the samples to estimate θ and hence f. Complications arise, however, owing to the fact that the distribution of v is not independent of that of Σ_1/Σ_2. A discussion of this point is beyond the scope of this paper.

7. *The fiducial test of R. A. Fisher.* I have so far considered only two of many criteria which may be proposed to test the hypothesis that $\alpha_1 = \alpha_2$. I have moreover been concerned with one aspect only of the tests based upon these two criteria, viz. whether they control satisfactorily the risk of rejecting the hypothesis when it is actually true. A test which does control this risk is termed in the present paper either a "valid" or an "unbiased" test. The special sense in which these terms have been used should be noted, for a valid test is not necessarily a good test nor vice versa. In the present section I propose to discuss the fiducial test suggested by R. A. Fisher from the single point of view only of how far it is valid, in the sense defined above, for all values of the ratio $\theta = \sigma_1^2/\sigma_2^2$.

The manner of developing this fiducial test is as follows†: let

$$\left.\begin{aligned} &s_1 = \sqrt{\frac{\Sigma_1}{n_1(n_1-1)}}; \quad s_2 = \sqrt{\frac{\Sigma_2}{n_2(n_2-1)}};\\ &d = (\bar{x}_1 - \bar{x}_2); \quad \delta = (\alpha_1 - \alpha_2); \quad \epsilon = (\delta - d);\\ &t_1 = \frac{(\bar{x}_1 - \alpha_1)}{s_1}; \quad t_2 = \frac{(\bar{x}_2 - \alpha_2)}{s_2}. \end{aligned}\right\} \qquad \ldots\ldots(15)$$

From (15) we obtain

$$\epsilon = (\delta - d) = s_2 t_2 - s_1 t_1. \qquad \ldots\ldots(16)$$

The fiducial distribution of δ is taken by Fisher to be the distribution obtained from (16) by treating d, s_1 and s_2 *formally* as fixed and allowing t_1 and t_2 to be distributed independently as t with degrees of freedom $(n_1 - 1)$ and $(n_2 - 1)$ respectively.

Now if A_1 and A_2 are any *constants* the distribution of $(A_2 t_2 - A_1 t_1)/\sqrt{A_1^2 + A_2^2}$ clearly depends only on A_1/A_2, n_1 and n_2. Hence we can theoretically determine

* See J. Neyman's paper, "Statistical Problems in Agricultural Experimentation". *J. Roy. Statist. Soc.* Supplement, II, No. 2 (1935), pp. 130–6. The sensitivity of t criteria to real population differences was seen to depend on f in a pronounced fashion only when f was very small (say < 6). In the present case the increase in sensitivity of u (with $f = 18$) over v (with $f = 6{\cdot}89$) will not be large.

† R. A. Fisher, *Ann. Eugen.* VI, Part IV (1935), p. 396.

a function $F(A_1/A_2, n_1, n_2, \gamma)$ such that the probability is γ that the inequality

$$|A_2t_2 - A_1t_1| > \sqrt{A_1^2+A_2^2}\, F(A_1/A_2, n_1, n_2, \gamma) \qquad \text{......(17)}$$

is satisfied. The corresponding statement in terms of fiducial probability is that the inequality

$$|\delta - d| > \sqrt{s_1^2+s_2^2}\, F(s_1/s_2, n_1, n_2, \gamma) \qquad \text{......(18)}$$

is satisfied with fiducial probability γ. The corresponding fiducial test of the hypothesis $\delta = 0$, which Fisher has suggested, consists in rejecting the hypothesis if

$$\left|\frac{d}{\sqrt{s_1^2+s_2^2}}\right| > F(s_1/s_2, n_1, n_2, \gamma), \qquad \text{......(19)}$$

where γ is now the level of significance. However, in repeated sampling from fixed populations π_1 and π_2 with $\alpha_1 = \alpha_2$, the repeated application of this test will not, in general, lead to rejection in the prescribed proportion, γ, of cases. Although the inequality (17), which can be written

$$\left|\frac{A_2(\bar{x}_2-\alpha_2)}{s_2} - \frac{A_1(\bar{x}_1-\alpha_1)}{s_1}\right| > \sqrt{A_1^2+A_2^2}\, F(A_1/A_2, n_1, n_2, \gamma) \qquad \text{......(20)}$$

is satisfied with probability γ whatever constant values are assigned to A_1 and A_2, this does not imply that the inequality (18) which can be written

$$\left|\frac{s_2(\bar{x}_2-\alpha_2)}{s_2} - \frac{s_1(\bar{x}_1-\alpha_1)}{s_1}\right| > \sqrt{s_1^2+s_2^2}\, F(s_1/s_2, n_1, n_2, \gamma) \qquad \text{......(21)}$$

will also be satisfied with the same probability γ.

It may be noted that in the case $n_1 = n_2 = 2$, Fisher has himself drawn attention to the difference between (20) and (21). As he has shown for that case*, the fiducial test involves the rejection of the hypothesis $\delta = \alpha_1 - \alpha_2 = 0$ at a 5% level of significance when a certain criterion T, which is a function of the sample observations only, numerically exceeds 12·7062. He shows also, however, that if we are sampling from fixed normal populations π_1 and π_2 for which $\delta = 0$, the probability of $|T|$ exceeding 12·7062 is only equal to 0·05 if $\theta = \sigma_1^2/\sigma_2^2 = 0$ or ∞; in general the probability will be less than 0·05.

That the statement (18) may be associated with a specially defined measure of fiducial probability which could be used by the experimenter as a guide in deciding whether to reject the hypothesis $\delta = 0$ is quite possible. But it seems to me important to make clear that the rule of the test involved in (19), if applied to repeated samples taken from fixed normal populations π_1 and π_2, would not lead in the long run on a proportion γ of occasions to the rejection of the hypothesis $\delta = 0$, when it is true, whatever be the value θ.

* *Ann. Eugen.* VII, Part IV (1937), p. 374.

8. *Exact tests.** Dr Bartlett has pointed out to me that the exact test which he gives for the case $n_1 = n_2 = 2$, is capable of easy generalization. For instance, if $n_1 = n_2 = n > 2$, let $l_{11}, l_{12}, \ldots l_{1,n-1}$ be $n-1$ linear functions of the observations in the first sample: let the l's be orthogonal to each other and to $\bar{x}_1$: further, let them all have expectation zero and standard deviation σ_1. Linear functions satisfying these conditions can always be defined. Similarly define $l_{21}, l_{22}, \ldots l_{2,n-1}$ for the second sample, these having standard deviation σ_2. Then $\sqrt{n}\,(\bar{x}_1 - \bar{x}_2)$ divided by $\sqrt{\left\{\sum_{i=1}^{n-1} (l_{1i} + l_{2i})^2\right\} \Big/ (n-1)}$ will be a criterion distributed as t with $(n-1)$ degrees of freedom, whatever the ratio σ_1^2/σ_2^2. Clearly a like test can be evolved when $n_1 \neq n_2$, the degrees of freedom of the corresponding t, then being one less than the smaller of n_1 and n_2. Bartlett would not advocate the use of this test in practice for the reason that it is not more efficient than using an inexact test based on v, and expressing the significance level of the sample as lying between two limits. The number of degrees of freedom for the criterion defined above is $n_1 - 1$ (if $n_1 < n_2$), whereas the effective number of degrees of freedom for v is never less than $(n_1 - 1)$ and may be as much as $(n_1 + n_2 - 2)$.

While on the subject of exact tests, it is of interest to note that other criteria of the form $(\bar{x}_1 - \bar{x}_2)/\sqrt{d\Sigma_1 + e\Sigma_2}$ may be less dependent on σ_1^2/σ_2^2 even than v. For instance, if n_1 and n_2 are both > 3 we might expect

$$z = (\bar{x}_1 - \bar{x}_2) \Big/ \sqrt{\frac{\Sigma_1}{n_1(n_1-3)} + \frac{\Sigma_2}{n_2(n_2-3)}}$$

to be such a criterion. The reason for taking these particular values of d and e is that they give to σ_z^2 the same value both when $\theta = \sigma_1^2/\sigma_2^2 = 0$ and when $\theta = \infty$. The curve (c) in Fig. 2 shows the dependence of z on θ. Arranging the test so that the probability of rejection of the hypothesis $\alpha_1 = \alpha_2$ is 0·05 when $\theta = 1$, it is seen that no matter what θ, the probability of rejection departs from 0·05 less for z than for either of the criteria u and v. It is not proposed in the present paper to discuss whether tests such as these are of practical value.

9. *Summary.* Three tests of the hypothesis that the means of two normal populations are equal have been considered in some detail. The object has been to study how closely each of these controls the risk of rejecting the hypothesis when it is actually true. None of the tests was exact in the sense that it would control this risk precisely, whatever the unknown ratio θ of the variances of the two populations.

The first criterion u, which is the best criterion when it is known that θ is unity, can under certain circumstances be seriously biased when $\theta \neq 1$.

* By an exact test is meant one depending on a known probability distribution; that is, independent of irrelevant unknown parameters (e.g. in the present case independent of $\theta = \sigma_1^2/\sigma_2^2$). See, for instance, M. S. Bartlett, *Proc. Roy. Soc.* A, CLX (1937), p. 271.

The second criterion v, which employs separate estimates of the unknown variances of the two populations, was seen to be very much less liable to bias. Unless therefore it is definitely known that $\theta = 1$, the general use of v rather than u is worth serious consideration.

I agree with M. S. Bartlett's criticism of the third test, which has been put forward from considerations of fiducial probability. The bias of this test depends on θ, but I have not considered the relationship in any detail.

NOTE ON AN APPROXIMATION USED BY B. L. WELCH

By Elizabeth Tanburn, B.A.

In the preceding paper on the "Significance of the Difference between Two Means", B. L. Welch has considered two criteria, viz.

$$u = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\dfrac{\Sigma_1 + \Sigma_2}{(n_1+n_2-2)}\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}}; \quad v = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\dfrac{\Sigma_1}{n_1(n_1-1)}+\dfrac{\Sigma_2}{n_2(n_2-1)}}}.$$

He discusses the distribution of these in the case where the means α_1 and α_2 of the normal populations sampled are equal, but where the standard deviations σ_1 and σ_2 are not necessarily equal. He shows that u and v will be distributed approximately as ct_f, where c is a constant and t_f is distributed as "Student's" t having f degrees of freedom; c and f which are functions of n_1, n_2 and $\theta = \sigma_1^2/\sigma_2^2$ are given by equations (8) and (9) of p. 353 above.

The present writer has been studying the same problem both theoretically and also by means of practical sampling experiments and the results of this investigation will shortly be published. It may be of interest, however, to note here some points which have bearing on the approximation Welch has used. The approximation was made by fitting the quantities under the square roots in the criteria by Pearson Type III curves. The fitting was performed by making the Type III curves have the correct first two moments. This was a convenient method, but, of course, not the only one. We might for instance write $u \doteq ct_f$ and choose c and f so that the μ_2 and β_2 of ct_f are the same as the true μ_2 and β_2 of u, in other words represent u (and v) by a Pearson Type VII curve having the correct 2nd and 4th moments. In general, the objection to this is the more complicated form of the moments of u. (Also for very small samples these moments become infinite.) I have, however, obtained μ_2 and β_2 for u and v in one particular instance, viz. $n_1 = 5$, $n_2 = 15$, $\theta = \sigma_1^2/\sigma_2^2 = 0{\cdot}25$. They are given in the first line of Table I. Let us now compare them with the moments of Welch's approximation obtained by taking u (or v) $= ct_f$. For this we have

$$\mu_2 = \frac{c^2 f}{(f-2)}; \quad \beta_2 = \frac{3(f-2)}{(f-4)}.$$

TABLE I

	$\mu_2(u)$	$\beta_2(u)$	$\mu_2(v)$	$\beta_2(v)$
True moments	0·6002	3·493	1·1512	3·503
Moments of Welch's approximation	0·6011	3·509	1·1608	3 575

For u equation (8) gives $f = 15{\cdot}79$, $c^2 = 0\,5250$. For v equation (9) gives $f = 14{\cdot}44$, $c^2 = 1$. Substituting these we obtain the values given in the second line of Table I. The agreement with the true values is close enough to indicate that no great difference in the values of c and f would have occurred if Welch had used the moment method to represent u (or v) by a Type VII ("Student") curve rather than to represent the square of the denominator of u (or v) by a Type III (χ^2) curve.

A further comparison, which is worth making, is that between the approximate theoretical distribution of u and the actual distribution of 500 values of u which were obtained in a sampling experiment using Tippett's Random Numbers ($n_1 = 5$, $n_2 = 15$, $\theta = 0{\cdot}25$ as before). In the second line of Table II are shown the numbers of these sample u's whose absolute values lay between the limits given

TABLE II

$\lvert u \rvert$	0 000–0·362	0·362–0·500	0·500–0·627	0·627–0·777	0·777–0·969	0 969–1·266	1·266–1·538	1·538–1·874	> 1·874
Frequency of samples	252	50	57	44	53	26	12	1	5
Approx. theoretical chances	0 50	0·10	0·10	0·10	0·10	0·05	0·03	0·01	0·01
Approx. expectations	250	50	50	50	50	25	15	5	5

in the first line. These limits are so chosen that, if the representation, $u = ct_f$, were exact (f being 15·79 and c^2 being 0·5250), then the true theoretical chances of u falling in the ranges would be those given in the third line of the table. The corresponding expectations for 500 samples are given in the last line. The sampling results are seen to agree very well with the theoretical distribution as it has been approximated.

COMPARISON BETWEEN BALANCED AND RANDOM ARRANGEMENTS OF FIELD PLOTS

By "STUDENT"

[With very deep regret the Editorial Committee has to report the death on 16 October 1937, of Mr W. S. Gosset, whose scientific contributions under the pseudonym of "Student" are well known to all statisticians. It is hoped to include some account of his life and work in the next issue of the *Journal*.

Mr Gosset had been working at the following paper during the past summer, and a fortnight before his death had discussed the draft, which is printed below, with Dr J. Neyman and Prof. E. S. Pearson. It was then agreed that certain points in sections 2 and 3 needed clarification and Mr Gosset proposed to undertake this work himself; unfortunately this final revision was never completed. Dr Neyman and Prof. Pearson have therefore added in a separate Note (pp. 380–88 below) some comments, for which they take full responsibility, regarding the points on which they know Mr Gosset had intended to enlarge.—Ed.]

In a paper read before the agricultural and industrial section of the Royal Statistical Society* I ventured to point out that the advantages of artificial randomization are usually offset by an increased error when compared with balanced arrangements. Prof. Fisher does not agree and has written a paper to test the difference of opinion that there is between us.†

In this paper I propose to set out as clearly as I can just what is this difference of opinion.

Next I propose to show that the conclusions of Prof. Fisher's paper all follow firstly from his having made use of a method of calculating the error of the "systematic" arrangements which I showed fourteen years ago would lead to just the misleading conclusions which he has found, and secondly to his not having compared like with like.

Thirdly, I will show that if he had not fallen into these pitfalls he would have been able to show that in the case which he took, a balanced arrangement does in fact give a slightly smaller error than his randomized one.

Fourthly, I will describe just what is to be expected when balanced arrangements are compared with random,‡ viz. that when the variance due to treatment is low compared with the error of the experiment, fewer significant results are obtained than with random arrangements, but when the variance due to treatment is high more significant results are obtained with balanced arrangements.

* "Co-operation in large-scale experiments," W. S. Gosset, Supplement to *J. roy. Statist. Soc.* III (1936), 115–22.

† "A test of the supposed precision of systematic arrangements," Barbacki and Fisher, *Ann. Eugen.* VII (1936), 189–93.

‡ Note that an arrangement can be both balanced and random and where this is practicable the aims of Prof. Fisher and myself are both satisfied.

Lastly, I will give in an appendix the results of some testing of balanced versus random arrangements on uniformity trials by Mr A. W. Hudson of Massey College, N.Z.

§ 1. The effect of lack of randomness on bias

It is almost invariably necessary, when applying mathematics to practical affairs, to replace the actual conditions by a set of simpler approximations with which the mathematics are capable of dealing, and mathematical statistics are no exception to this rule.

For example, the analysis of variance which is generally used to determine the error of agricultural experiments requires three assumptions to be made before we can apply the method strictly:

(1) The systems concerned are to have normal variation.

(2) The variances of like things should be equal.

(3) The sampling should be random.

(1) If, as is usual, the variation is not normal our argument will not be impaired unless the number of replications is very small, when departure from normality introduces an added uncertainty to the estimation both of mean and perhaps even more of variance.

(2) If, as often happens, the variances are not equal, as for example when we are pooling the variances of the yields of barleys which react differently to soils of different fertility, we shall not in general invalidate our conclusions appreciably, though in extreme cases attention should be paid to this source of error.

(3) If, however, the sampling be not random, there are such possibilities of drawing false conclusions that Prof. Fisher has introduced a system of artificial randomizing to ensure that the third condition is satisfied and brands all other systems invalid.

Nevertheless, it is possible, by balancing sources of error which would otherwise lead to bias, to obtain arrangements of greater precision which are nevertheless effectively random, by which I mean that the departure from randomness is only liable to affect our conclusions to the same sort of extent as do departures from normality or inequality of variances.

Lack of randomness can affect either the mean or the variance, and it is the first of these which is apt to lead to invalid conclusions. Thus Mr Yates has shown that it is practically impossible for anyone to select shoots of corn of average length by eye, and in fact none of the senses can be trusted to behave without bias. Those of taste or smell are peculiarly liable, and if comparisons are to be made it is necessary to avoid giving the least inkling of the order in which the samples are to be presented, in fact it is better to let it be known that it is a random order. In some cases the only way of avoiding bias is to withhold all knowledge of the object of the investigation from those taking part, though unfortunately this engenders a lack of interest in the proceedings.

Again, a promising experiment in nutrition was ruined by departure from randomness when the schoolmasters were allowed to adjust the supposed uneven effects of a chance selection of subjects for the Lanarkshire Milk Experiment, and in doing so managed to select, doubtless from the most humane motives, 10,000 children to receive milk who were significantly lighter and shorter than the 10,000 "controls" who did not.

In agricultural experiments there are obvious possibilities of bias affecting the mean in badly arranged experiments, for it is usual to find "fertility slopes" in most "uniformity" experiments, i.e. when an apparently uniform field is harvested in small-sized plots it is usual to find that the yield is higher in some parts than in others and tends to change more or less gradually from one place to another. Hence if plots of one variety are sited, whether systematically or by chance, nearer to one end of the experimental area than to the other, the mean is likely to be biased.

To take the simplest case of two varieties or treatments, the layouts

$$A\ B\ A\ B\ A\ B\ A\ B \text{ (systematic)}$$

and

$$A\ B\ A\ A\ B\ A\ B\ B \text{ (random)}$$

will both favour B if the field is more fertile on the right than on the left hand, the second rather more than the first.

On the other hand the layout

$$A\ B\ B\ A\ A\ B\ B\ A$$

is balanced with regard to a simple "linear" fertility slope, and the mean of neither A nor B will be biased except by departure from linearity.

It is, of course, possible to imagine particular variations in soil fertility which will bias the means of plots arranged in this manner, but with one exception they are of the same nature and lead to the same sort of bias—but usually to a smaller extent—as occurs with artificially randomized layouts.

The one exception is a periodic wave of fertility due to previous cultivations which happens to coincide in period with the width of an odd integral number of quartets, a not particularly likely occurrence.

Such layouts as $ABBA$ are termed balanced, and any number of treatments may be set in a balanced layout, as, for example, in the Latin square which is not only balanced but random as well, "thus conforming to all the principles of allowed witchcraft".

It is reasonable to expect that balanced layouts will on the whole be successful and that the mean will be less biased than in random, and this expectation is illustrated by some experimental sampling carried out by Mr A. W. Hudson of Massey College, N.Z., who tested balanced and random blocks against one another on three different uniformity trials. His results are given in the Appendix, and all

that need be said here is that in fifteen experiments the balanced layouts showed slightly more bias in three and less in twelve, the reduction of bias being very considerable in some of the twelve.*

And this brings me to a question which has often interested me. Suppose there are two treatments to be randomized—I take two for simplicity only—and suppose that by the luck of the draw they come to be arranged in a very *un*balanced manner, say *AAAABBBB*· is it seriously contended that the risk should be accepted of spoiling the experiment owing to the bias which will affect the mean if there is the usual fertility slope? For, as will be shown later, not only will the mean be biased, but the apparent precision will tend to be high, and misleading conclusions drawn much more often than the 1 or 5 % of the tables. It is of course perfectly true that *in the long run*, taking all possible arrangements, exactly as many misleading conclusions will be drawn as are allowed for in the tables, and anyone prepared to spend a blameless life in repeating an experiment would doubtless confirm this; nevertheless it would be pedantic to continue with an arrangement of plots known beforehand to be likely to lead to a misleading conclusion.

Let us suppose therefore—as indeed it is rumoured—that common sense prevails and chance is invoked a second time and that such an arrangement as *BBABBAAA* is offered; is this to be accepted? It is more likely to give a biased mean than *BABABABA*, but then of course it is random!

And if this is not to be used, how about *BBABABAA*? In short, there is a dilemma—either you must occasionally make experiments which you know beforehand are likely to give misleading results or you must give up the strict applicability of the tables; assuming the latter choice, why not avoid as many misleading results as possible by balancing the arrangements? And this, to do Prof. Fisher justice, is the direction towards which he is tending; in his paper with Dr Barbacki he treats for the first time of "randomized sandwiches" to which the objection is, not an appreciable increase of error, but the practical difficulty of working them.

To sum up, lack of randomness may be a source of serious blunders to careless or ignorant experimenters, but when, as is usual, there is a fertility slope, balanced arrangements tend to give mean values of higher precision compared with artificial arrangements.

Next, what is the effect of lack of randomness on the variance?

In a later section I will show that since in the "null" case, i.e. when no real treatment differences exist, the aggregate variance due to "treatments" and residual error is constant for all arrangements of treatments in the blocks, those with low actual error necessarily give high calculated values for the error and vice versa, the calculated error, however, varying much less than the actual in

* Mr Borden, of Hawaiian Sugar Planters' Association, Hawaii, has obtained similar results in similar experiments, and I have no doubt that this will always tend to happen.

ordinary experiments owing to the larger number of degrees of freedom of the residual error.

This, of course, has nothing to do with the origin of the experiment whether randomized or not.

If, however, the arrangement is "randomized" one can—*before the draw*—state accurately, subject to normality, etc., what the chance of getting any particular partition of variance between "treatment" and "residual error" will be in the "null" case. After the draw, when one particular arrangement has been chosen, it is often possible to be sure that the chance has changed in one direction or another without, however, being able to define exactly what it is.* In particular, balanced arrangements tend to have lower actual errors and higher calculated errors than would be expected by chance before a random selection is made, and this is so even if a degree of freedom is allocated to fertility slope, owing to the departure of the "slope" from linearity.

The consequence is that balanced arrangements more often fail to describe small departures from the "null" hypothesis as significant than do random, though they make up for this by ascribing significance more often when the differences are large.

Thus such departures from the "null" hypothesis as are found to be significant by balanced are likely to be larger than those found by randomized arrangements, and in particular those discovered in the "null" case itself—5 or 1 % as the case may be—tend to disappear altogether with balanced arrangements.

It will be seen then that the difference between Prof. Fisher and myself is not a matter of mathematics—heaven forbid—but of opinion. He holds that balanced arrangements may or may not lead to biased means according to the lie of the ground, but that in any case the value obtained for the error is so misleading that conclusions drawn are not valid, while I maintain that these arrangements tend to reduce the bias due to soil heterogeneity and that so far from the conclusions not being valid they are actually less likely to be erroneous than those drawn from artificially randomized arrangements. Further, that in the really important agricultural experiments which are carried out at more than one centre—and it was of these that I was speaking—the very slight disadvantage that an occasional result at an individual station may not be recognized as significant owing to over-estimation of the error at that station is more than offset by the greater precision of the experiment as a whole.

* This is analogous to the use of a life table to give the expectation of life. Thus the expectation of life of an Englishman of 40 can be referred to an appropriate table, but when we particularize the Englishman of 40 as a tin-miner or an agricultural labourer we know that the expectation is lower or higher than that given in the table without perhaps knowing very exactly by how much.

§ 2. Barbacki and Fisher

Such being our opinions, based in each case on *a priori* argument, Prof. Fisher rightly decided to put the matter to the test by assigning imaginary treatments to plots of which the yield had been determined in a uniformity experiment both on a random and on a balanced system, and published a paper,* of which he gives the following summary:

"1. This enquiry was carried out to test the truth of the opinion expressed by 'Student' that randomization achieves its object 'usually at the expense of increasing the variability when compared with balanced arrangements', and that one of the means available to experimenters of reducing the error is by adopting a 'regular balanced arrangement'.

"2. Using an extensive uniformity test it is found that the arrangements randomizing either pairs or sandwiches of half drill strips give smaller errors than the systematic arrangement advocated as more precise.

"3. As a consequence experimenters using the systematic arrangements systematically underestimate their errors.

"4. The error estimated from a systematic arrangement is ambiguous, and the experimenter has an arbitrary choice between several widely different estimates.

"5. Owing to the failure to furnish a valid estimate of error, 'Student's' test of significance is not approximately correct for systematic arrangements."

The particular arrangement which Prof. Fisher intended to test was the Half-Drill Strip† introduced by Dr Beaven some 14 years ago and widely used since then, but unfortunately half-drill strips are too large to lend themselves easily to testing on ordinary uniformity trials, and although Prof. Fisher has laid out eight pairs of half-drill strips on his uniformity trials he has not in fact compared them with a corresponding random arrangement but has cut them up transversely into 5-yard lengths and has compared the actual error of the large half drill strips with that calculated from the randomized‡ sheaf weights of which they are composed.

Now it happens that Dr Beaven had originally proposed to calculate the error of the half-drill strip from sheaf weights of this kind, and that I pointed out in this *Journal* 13 years ago§ that since such "sheaf weights" may be positively correlated such a method of calculating the error is fallacious.

This method of calculating the error has, of course, nothing to do with balanced arrangements, except that it was proposed by Dr Beaven, the author

* "A test of the supposed precision of systematic arrangements", Barbacki and Fisher, *Ann. Eugen.* VII, 189–93.

† Prof. Fisher prefers to call this the "Split Drill" Method, but though I agree that the name is more descriptive it is a pity to confuse the matter by a change of name after all these years. More particularly is it confusing to transfer the name "Half-Drill Strip" to small portions of the original half-drill strip as he has done, and I have called them by Dr Beaven's name of "Sheaf Weights".

‡ *Not very much randomized*; he compares corresponding pairs just as anyone else would.

§ "On testing varieties of cereals", *Biometrika*, xv (1923), 271–93.

of the half-drill strip; it might just as well be applied to random arrangements, as, for example, the "randomized pairs" of Prof. Fisher's experiment, each of which was actually harvested in six separate drills from which the error could have been equally erroneously calculated.

Prof. Fisher has therefore calculated the error of the half-drill strip by a method which I showed 13 years ago would be likely to give a fallaciously low value, and quite rightly has not used this method to calculate the error of his "randomized pairs": it is entirely due to this that he can draw conclusion (2) of his summary.

From this single fallacious conclusion he boldly generalizes to reach conclusion (3) which, as was shown by O. Tedin whom he quotes, is directly at variance with the facts. Conclusion (5) also follows solely from Prof. Fisher's faulty method and not from the balanced arrangement.

When the paper appeared I wrote a letter to *Nature* pointing this out, and that the actual error of the half-drill strip aggregate was in good conformity with that calculated from the weights of the whole strips.

In answering me Prof. Fisher replied that in that case the error of the "randomized sheaf weights" was so much smaller than that of half-drill strips that eleven times the area would have to be used to reduce the error of half-drill strips to that of "randomized sheaf weights" and further repeating his conclusion (4) with which I shall deal later.

Now one of the things that was noticed when uniformity trials first began was that the same piece of land laid out in large plots gave a very much larger error than if subdivided into small plots, and since half-drill strips were in this trial twelve times as large as "sheaf weights", Prof. Fisher's conclusion naturally follows since he is not comparing like with like.

Yet even so, those who have actually had to carry out agricultural experiments might very well prefer to work eleven times the area with ordinary agricultural methods and tools than have to sow and harvest 192 "randomized sheaf weights", if indeed that could be done at all under ordinary weather conditions.

Nevertheless, it is a fact that the error of this particular set of half-drill strips is unusually large. This arises partly because the number of repetitions is low but chiefly from the fact that the uniformity trial which Prof. Fisher chose to illustrate his argument showed a rather unusual feature due to faulty technique.

An examination of the original drills which were condensed to form the half-drill strips shows a periodicity, the averages of each eighth drill being for fifteen repetitions:

6739 7200 7839 6795 6689 7478 6897 6697

These variations are obviously not due to chance (for instance, the third drill gave the highest yield in twelve of the sets of eight and second highest in the other three) and are doubtless connected with some defect in the seed drill, probably the

tines were not evenly spaced, and this could possibly have been detected had it occurred to Mr Wiebe to examine the working of the drill before sowing.

The result is that since six of the eight drills were added up to form a "half-drill strip", then one drill omitted, and then another six, and so on, there was a periodic variation in fertility not coinciding in period with the width of the half-drill strip, and this, as I pointed out in the Appendix to my Royal Statistical Society paper, increases the calculated error but does not bias the mean.

For the same reason the correlation between the corresponding sheaf weights is very much higher than would usually be the case and full scope is thereby given to Prof. Fisher's faulty method of calculating the error.

Let us now deal with Prof. Fisher's fourth conclusion: "The error estimated from a systematic arrangement is ambiguous and the experimenter has an arbitrary choice between several widely different estimates."

We may observe in passing that this is another instance of Prof. Fisher's passion for generalizing on somewhat narrow foundations, for the possibility which he refers to is peculiar to the half-drill strip arrangement.

In the half-drill strip, however, it is possible either to calculate the error from such aggregates as $ABBA$ which I termed sandwiches in my paper to this *Journal* or from the separate parts of such aggregates, AB and BA, termed "pairs" by Prof. Fisher.

Of these the former is clearly the better if only there is a sufficient number of replications to give a good estimate of the error. As this is unusual it is generally best to give a degree of freedom to the fertility slope and calculate the error from "pairs".

Admittedly this tends to overestimate the error with the sort of results obtained in § 4. Faced with this choice, I personally choose the method which is most likely to be profitable when designing the experiment rather than use Prof. Fisher's system of *a posteriori* choice* which has always seemed to me to savour rather too much of "heads I win, tails you lose".

§ 3. A properly balanced arrangement

It appears then that Prof. Fisher's paper is altogether irrelevant to the question at issue, but in order that Dr Barbacki's work may not be wholly wasted we can make a calculation of the error of a properly balanced arrangement of plots of the same size as the "randomized sandwiches" of which he has calculated the error.

For it will be noticed that Prof. Fisher's "systematic" arrangement, though "balanced" as "half-drill strips", is not so when regarded as a number of "sheaf weights": lateral balance is necessary.

The obvious layout is therefore to have the $ABBA$ arrangement in both directions.

* *Statistical Methods for Research Workers*, § 24.1 (5th ed.), p. 125.

Thus:

A B B A A B B A A B		*A A A A A A A A*
B A A B B A A B B A		*B B B B B B B B*
B A A B B A A B B A	etc. instead of:	*B B B B B B B B* etc.
A B B A A B B A A B		*A A A A A A A A*
A B B A A B B A A B		*A A A A A A A A*
B A A B B A A B B A		*B B B B B B B B*
etc.		etc.

This is merely a chessboard with fringes, each square being divided at harvest into four. The "squares" should be long and narrow, to gain the advantage of contiguity, and the comparisons should be made between adjacent long subplots of the different varieties. I have not seen this rather obvious arrangement mentioned before; it is admittedly no more suited for agricultural work than "randomized sandwiches", but it might be used in horticultural work, where the reduced "borders" would be of advantage, or for pot culture.

In this case we can start from Dr Fisher's Table II by reversing the signs of columns (ii), (iii), (vi), (vii), (x) and (xi) and calculate the error from an analysis of variance as follows:*

Variance due to	Degrees of freedom	Sum of squares of "split drill" differences
Longitudinal fertility slopes	12	887,171
Lateral fertility slopes	8	4,508,506
Varietal difference	1	2,741
Residual errors	75	3,988,681
Total	96	9,387,099

The difference between A and B is thus 513 g. and the S.D. of this difference 2259, as compared with 2353 calculated from "random sandwiches".

Thus, as we should expect, the difference is comfortably within the S.D., and the S.D. a little below that calculated from "randomized sandwiches", itself a partially balanced arrangement though random.

We see then that if a properly balanced arrangement is put down on the uniformity experiment of Dr Fisher's choice the error is found to be, as usual, less than his random arrangement, though not by much since "sandwiches" are themselves balanced.

* See note regarding this analysis on pp. 384–88 below. [ED.]

§ 4. The effect of "balancing" on the "validity" of conclusions

From *a priori* considerations—and Mr Hudson's and Mr Borden's experiments are in accordance with this expectation—it seems fairly certain (i) that "balancing" has no tendency to bias the mean, and (ii) that when there is a "fertility slope"—or anything corresponding to it, e.g. a time effect—the result will be to increase the apparent error but to decrease the real error. What effect has this on the "validity" of conclusions drawn from balanced experiments?

(i) *The case of blocks, randomized or balanced, judged by the z test*

Let us take the case of four treatments in six blocks giving fifteen degrees of freedom to the residual error and three for treatments, and let us suppose the arrangement put down on a uniformity trial.

Then, once the plots and blocks are marked out, the "total sum of squares" and the "sum of squares due to blocks" are fixed; the difference between these represents in all cases the eighteen degrees of freedom due to treatments and residual error, but will be divided between the two in different proportions according to the chosen arrangement of the treatments in the blocks. If the arrangement is random the frequency of any particular ratio is known to follow the z distribution, and owing to the skewness of this there will more often than not be a lower variance of the treatments with three degrees of freedom than of the residuals with fifteen.

If the arrangement is not random the frequencies will not follow the z distribution, e.g. with regular unbalanced arrangements the variance "due to treatment" will tend to be high compared with that of "residual error", while with regular balanced arrangements the reverse is the case. It will therefore be of interest to see what happens when a real "variance due to treatment" is imposed on uniformity trials which give ratios at different points of the z scale.

Thus it may be convenient to take as norm those uniformity trials which have the same variance for "means of treatments" as that calculated from the residuals and let this variance be σ^2. Then another set of trials may be considered of which the means have a variance of $0{\cdot}5\sigma^2$ and consequently a variance of "residual error" of $1{\cdot}1\sigma^2$, since $15 \times 1{\cdot}1 + 3 \times 0{\cdot}5 = 18$. This set may be taken to represent the tendency of balanced arrangements to produce low variance "due to treatment". A third set representing "unbalanced" arrangements may be taken with a means variance $1{\cdot}5\sigma^2$ and a variance calculated from residuals of $0{\cdot}9\sigma^2$.

All three of these occur, of course, in their proper proportions in random trials and are none of them uncommon. They are merely taken here as types.

In what follows I shall for convenience term the variance of means the *actual*

variance of error, σ_e^2, and the variance calculated from residuals the *calculated* variance of error.

Now suppose that a real variance due to treatment—measured without error, σ_T^2—be superposed upon the uniformity experiment. Then the calculated variance of error will be unaffected and the observed variance due to treatments will be $\sigma_T^2+\sigma_e^2+2r_{Te}\sigma_T\sigma_e$ and, since T and e are independent, the distribution of the observed variance can be calculated from the known distribution of r when there is no correlation, which in this case of four treatments is uniform between $+1$ and -1.

From this we can determine the probability that any given σ_T^2, superposed on any particular arrangement, will be deemed "significant" when compared with the corresponding "calculated variance of error".

The results of such calculations are given in the following table, which gives the probability of exceeding the 5 % limit of significance, or if preferred can be read as the percentages of "significant" results.

Value of σ_T^2/σ^2	Probability of obtaining significant result		
	Actual variance of error		
	$1{\cdot}5\sigma^2$	$1{\cdot}0\sigma^2$	$0{\cdot}5\sigma^2$
	Limit of significance		
	$2{\cdot}96\sigma^2$	$3{\cdot}29\sigma^2$	$3{\cdot}62\sigma^2$
0·5	0·22	0	0
1·0	0·41	0·18	0
1·5	0·51	0·34	0·03
2·0	0·58	0·45	0·22
2·5	0·63	0·53	0·36
3·0	0·68	0·60	0·48
3·5	0·72	0·66	0 57
4·0	0·76	0·71	0·66
4·5	0·79	0·76	0·73
5·0	0·82	0·80	0·80
5·5	0·85	0·84	0·86
6·0	0·88	0·88	0·92
6·5	0·90	0·91	0·97
7·0	0·93	0·94	1·00
7·5	0·95	0·98	—
8·0	0·97	1·00	—
8·5	0·99	—	—
9·0	1·00	—	—

This table illustrates the fact that arrangements which give an actual error less than the calculated fail to give as many "significant" results as those which

give larger actual errors up to a real treatment variance of about five times the average residual variance, at which point about 20 % of the experiments still fail to show significance in each case. When the real treatment variance rises above this point, the smaller the actual error the *more* are the significant results.

It is perhaps rather invidious to decide below what value of the real treatment variance "significant" results are misleading, but in any case it is clear that the fault of the arrangements with low actual variance is not lack of validity. On the contrary, conclusions drawn from experiments giving significant results by such arrangements are *more* valid in the ordinary sense of that word.

These arrangements have so far been considered as having arisen in a random manner, but by using balanced arrangements the proportion of arrangements having actual low errors is increased, and hence conclusions arrived at from balanced arrangements are more, not less, valid.

Nevertheless, it is clear that if it is required to calculate the error from an experiment carried out at a single station it is advisable not only to balance the experiment but to allow for the error eliminated by allocating a degree of freedom to the fertility slope. Even so it is likely that the actual error will be less than the calculated and the conclusions more valid than they appear to be.

(ii) *The case of half-drill strips judged by the t test*

I showed in the appendix to my paper on Co-operative Experiments that it is usually advantageous to allot one degree of freedom to the fertility slope, and that since fertility slopes are not usually strictly linear there is a tendency for the calculated error to be larger than the actual error. Let us illustrate this in the case of experiments carried out on the scale adopted by the N.I.A.B., namely, with ten pairs of comparisons; this is of course rather a small scale, and of the nine degrees of freedom one is allocated to the fertility slope and eight to the residual error of comparing the two varieties.

In this case we are to vary, not the position of treatments on a given piece of ground, but the pieces of ground on which a half-drill strip of ten pairs is set and the "norm" which we shall take is the case where, owing to a particularly uniform fertility slope, the calculated and the actual error exactly correspond with the standard error σ.

With this we can compare a case where the variance of actual error is $0{\cdot}5\sigma^2$ and the calculated error therefore $\left(1+\frac{0{\cdot}5}{8}\right)\sigma^2 = 1{\cdot}062\sigma^2$, i.e. standard errors $0{\cdot}71\sigma$ and $1{\cdot}03\sigma$. A tendency in this direction is, as noted above, common, since fertility slopes are naturally not uniform; on the other hand, when the fertility slope is small random sampling may give us a case where the actual error is larger than the calculated, let us say standard errors of $1{\cdot}22\sigma$ and $0{\cdot}97\sigma$.

Then in the three cases we find from the t table that the 5 % significance point is for the "norm" $2{\cdot}30\sigma$, for the low actual error $2{\cdot}37\sigma$, and for the high actual

error $2{\cdot}23\sigma$, while the actual errors are distributed normally with S.E.'s. σ, $0{\cdot}707\sigma$ and $1{\cdot}22\sigma$ and the percentage of "significant" results, i.e. those above the significant point calculated above, can be readily determined for values of the real (i.e. measured without error) differences between the two "varieties", say $A-B$.

These are given in the following table.

Variance of calculated error	$0{\cdot}94\sigma^2$	$1{\cdot}0\sigma^2$	$1{\cdot}06\sigma^2$
Variance of actual error	$1{\cdot}5\sigma^2$	$1{\cdot}0\sigma^2$	$0{\cdot}5\sigma^2$
S.E. calculated	$0{\cdot}97\sigma$	$1{\cdot}0\sigma$	$1{\cdot}03\sigma$
S.E. actual	$1{\cdot}22\sigma$	$1{\cdot}0\sigma$	$0{\cdot}707\sigma$
Limit of significance	$2{\cdot}23\sigma$	$2{\cdot}30\sigma$	$2{\cdot}37\sigma$
Value of $\frac{A-B}{\sigma}$	Probability of significant results		
0	0·07	0·02	0
0·5	0·01 0·08	0·04	0
1·0	0·16	0·10	0·03
1·5	0·27	0·21	0·11
2·0	0·42	0·38	0·30
2·5	0·59	0·58	0·58
3·0	0·74	0·76	0·81
3·5	0·85	0·88	0·95
4·0	0·93	0·96	0·99
4·5	0·97	0·99	1·00
5·0	0·99	1·00	—
5·5	1·00	—	—

It will be noticed that in the left-hand column there are two probabilities given opposite 0·5, 0·01 that a negative significant result and 0·08 that a positive significant result will be obtained. Fortunately such a case is almost impossible unless of course "randomized pairs" were used instead of a half-drill strip. What we are concerned with in practice is something which tends towards the right-hand column which, as in the case of the balanced blocks, errs by failing to give significant results when the difference to be measured is small, but from a value of about 2·55—at which all produce significant results in 60 % of trials—gives a higher percentage than when the calculated and actual errors are equal.

It is clear, therefore, that in this case too, conclusions drawn from a balanced arrangement are not less but more valid than if the arrangement had been random.

The above tables rather emphasize the well-known paradox that it is just when the experimenter is congratulating himself on the unusual smallness of his experimental error—unusual, that is, for the type of experiment and number of replications—that he is most likely to be betrayed into drawing false conclusions: for the small calculated error indicates a large actual error, and this whether the

arrangement be random or balanced, though it is likely to occur more frequently in the random.

In conclusion, I should like to emphasize the fact that when using the phrase criticized by Prof. Fisher I was concerned with co-operative experiments carried out at a number of different places.

Such experiments, as indeed all agricultural experiments, are only of value in so far as the venue is representative of the conditions under which the results of the experiment are to be applied, and so the result at any single station is not of any particular importance in itself but only in its interaction with the results obtained at the other stations, for only so can its representative nature be established.

To take a simple case a variety trial may indicate that one wheat will do better than another in heavy but not in light soils; such a conclusion is more likely to follow from an experiment carried out with a low real error and a correspondingly high calculated error at the individual stations than if a low calculated error gave "significant" results sporadically.

It is therefore important that the results should be determined with as little real error as possible, and the calculated error at each station is superseded by the error of the experiment as a whole.

APPENDIX GIVING MR A. W. HUDSON'S COMPARISONS OF RANDOM AND REGULAR ARRANGEMENTS IN UNIFORMITY TRIALS

Mr Hudson's account of his procedure is as follows:

"(i) Four, five or six imaginary treatments were allocated according to which was the most suitable to the full utilization of the data.

"(ii) These were allocated to blocks in a regular-balanced fashion and then to the same blocks randomwise, using various numbers of 'units' per individual plot.

"The regular arrangements were balanced by using two or four series in which the treatments in the second and fourth series were in opposite order to those in the first and third, thus:

1, 2, 3, 4, 1, 2, 3, 4, etc.
4, 3, 2, 1, 4, 3, 2, 1, etc.
2, 1, 4, 3, 2, 1, 4, 3, etc.
3, 4, 1, 2, 3, 4, 1, 2, etc.

or alternatively, where the shape of the individual plot permitted, only a single series, thus:

etc., 2, 1, 4, 3, 2, 1 Middle 1, 2, 3, 4, 1, 2, 3, etc."

TABLE I

Data from Journal of Agricultural Science, Vol. IV, *Part* 2, 1911.
Mercer and Hall. Mangold Plots

Number of rows ... 20 } Total number of units 200, but only 160 used in first three.
Units per row ... 10 }

B./Tr.	R. × U.	G.M.	Random: Calculated S.E.	Random: Dev. of T.M. from G.M.	Random: Actual S.E.	Balanced: Calculated S.E.	Balanced: Dev. of T.M. from G.M.	Balanced: Actual S.E.
20/4	1 × 2	656·4	6·63	− 3·3 − 5·7 + 1·6 + 7·5	5·84	6·73	+ 4·4 − 1·0 − 1·7 − 1·7	2·95
10/4	2 × 2	1312·8	14·16	+10·2 − 4·1 −12·2 + 6·3	10·15	14·42	− 0·8 + 8·3 − 1·9 − 5·4	5·54
10/4	1 × 4	1312·8	16·40	+12·7 −18·0 + 7·5 − 2·0	13·48	16·61	+16·3 − 5·8 − 6·8 − 3·5	10·92
8/5	1 × 5	1642·9	21·62	−32·8 −20·0 +27·8 + 4·3 +20·5	25·9	22·92	+15·0 −16·2 − 5·5 +19·7 −13·2	16·4
4/5	2 × 5	3285·7	50·78	+55·0 −25 0 − 4·2 +43·0 −68·7	50·6	54·62	+15·3 −15·7 −53·2 + 9·3 +44·5	36·7

Table headings: B./Tr. Blocks (replications) and treatments.
R. × U. Size of plot, rows × units.
G.M. General mean of all plots.
Calculated S.E., i.e. of means of treatments by analysis of variance.
Dev. of T.M. from G.M. Deviation of treatment means from general means.
Actual S.E., i.e. calculated from previous column.

TABLE II

Data from Journal of Agricultural Research, Vol. XLIV, *No.* 8, *April* 1932.

F. R. Immer. Yields of sugar beet.

Number of rows	... ...	60
Units per row	... ...	10
Total number of units	... ...	600

B./Tr.	R. × U.	G.M.	Random			Balanced		
			Calculated S.E.	Dev. of T.M. from G.M.	Actual S.E.	Calculated S.E.	Dev. of T.M. from G.M.	Actual S.E.
20/6	1 × 5	255·9	3·28	+ 1·1 + 3·3 + 0·6 + 0·4 − 6·2 + 0·8	3·22	3·27	− 1·1 − 3·7 − 2·1 + 6·1 + 1·1 0	3·40
10/6	2 × 5	511·9	8·42	−11·0 +13·0 − 1·6 − 7·2 +13·0 − 6·2	10·5	8·52	+ 7·5 + 5·1 − 5·8 −17·3 + 2·6 + 7·8	9·8
10/6	1 × 10	511·9	8·04	− 6·1 + 7·4 − 4·4 + 9·5 − 0·4 − 6·0	6 90	8·11	− 6·2 − 1·7 +10·4 − 2·7 − 3·7 + 3·7	6·08
4/6	5 × 5	1279·7	23·68	+68 2 +40·5 −41·6 − 6·3 −73·3 +12·5	52·1*	37·87	+11·4 + 5·4 −17·0 + 2·1 − 5·4 + 3·6	10·0

* This is a "significant" result—beyond the 1 % level—and it is perhaps a little unfortunate that it should have occurred in a mere sample of 21. It has, however, been checked both by Mr Hudson and myself.

TABLE III

Data from Journal of Agricultural Science, Vol. XXII, *Part* 2, *April* 1932.
Kalankar. Potatoes.

Number of rows ... 96 Units per row ... 6 Total number of units ... 576

B./Tr.	R. × U	G.M	Random A			Random B			Balanced		
			Calc. S.E	Dev. of T.M. from G.M	Actual S.E.	Calc. S.E	Dev. of T.M. from G.M.	Actual S.E.	Calc. S.E.	Dev. of T.M. from G.M.	Actual S.E.
32/6	1 × 3	69 8	0 74	− 0·2 + 0·4 − 0·1 - 0·6 − 0 3 + 0·8	0·51	0 74	− 0·6 − 0·1 − 0·3 + 0·2 + 0·6 + 0 3	0 44	0·74	− 0·5 + 0 3 + 0·9 − 0·1 + 0·3 − 1·0	0·67
16/6	1 × 6	139 6	1·52	− 1 9 0 + 2·7 − 1·9 + 0·7 + 0·4	1·74	1 49	+ 1·1 + 1·8 − 3 1 + 1·2 + 1·1 − 2 1	2·05	1·55	− 1·4 + 1·3 + 0 9 + 0·7 0 − 1·5	1·20
16/6	2 × 3	139·6	2·16	+ 0·2 − 1·1 − 2·9 − 0·1 + 0·4 + 3·5	2 10	2·19	− 2·8 + 1·5 + 0·8 + 0·2 + 0 1 + 0·2	1 52	2 20	+ 0·2 − 0·3 − 2·1 − 0·5 + 2·0 + 0 8	1·38
8/6	2 × 6	279·2	5·35	+ 8·8 + 0 9 − 5·4 − 1·4 − 2·8 − 0·1	4·84	5·47	+ 0 8 + 6·0 + 2·1 − 3 8 − 1·1 − 4 1	3·83	5·56	+ 3·0 + 3 3 − 2·0 − 3·3 0 − 1·0	2·68
8/6	4 × 3	279 2	5·67	+ 1·3 + 6 1 + 1·1 − 3·9 − 3·5 − 1·1	3·71	5·58	+ 8·0 − 1·3 − 3·8 − 0·1 − 4·4 + 1·6	4·52	5·60	+ 2·4 − 4·4 − 2·2 − 2·7 − 0 9 + 7·7	4·42
4/6	8 × 3	558·4	33 3	−12·1 − 7·2 +41 0 +35·8 −19 0 −38 7	31·7	35·85	− 5·9 +39·0 +10·8 −16·9 −12·5 −14·6	21·6	37·76	− 3·3 −10·8 + 5·6 − 0·6 + 1·1 + 7·8	6·7

The above experimental work must not be taken as an attempt at a proof that balanced arrangements are likely to give a lower error than random unbalanced arrangements; that seems to me obvious, and it is for those who wish to disprove the obvious to obtain evidence in support of their eccentric opinions, but it does give an interesting illustration of what is likely to happen in practice, and I print it in the hope that it will help to clarify other people's ideas as it has mine.

NOTE ON SOME POINTS IN "STUDENT'S" PAPER ON "COMPARISON BETWEEN BALANCED AND RANDOM ARRANGEMENTS OF FIELD PLOTS"

By J. NEYMAN and E. S. PEARSON

During the summer of 1937 "Student" discussed the subject of this paper with one of us on several occasions. The paper was some months in preparation "not so much", he wrote, "owing to lack of time as to lack of inclination to controversy". He was particularly anxious, however, that it should contain as clear a statement as possible of his views on balanced and random arrangements, and when in the middle of September he sent us a final draft for comment, he asked us to make any suggestions for improvement we could think of. We told him in the first place that we felt that the reader who was not very familiar with the literature might find a little difficulty in following the points at issue between himself and Prof. Fisher. Secondly, on a smaller point, we suggested that he should explain somewhat more fully the analysis of variance carried out in his section 3.

These suggestions he welcomed, and a letter written four days before his death indicated that he was in the middle of adding to the paper some comments on these points. What form these additions would have taken we cannot tell with certainty, but we feel that it is right for us to add in a separate note a little fuller explanation of some of the points raised in sections 2 and 3 of the paper.

1. *The half-drill strip method.* In his paper "On Testing Varieties of Cereals" (*Biometrika*, vol. xv, 1923, p. 285 *et seq.*), "Student" described the half-drill strip method somewhat in the following terms:

When sowing, the seed box of the drill is divided into two across the middle, and the middle coulter put out of action. The seed of the two varieties, say A and B, is put in the seed box, one on each side of the division. Thus when sowing a drill strip, one half (i.e. six of seven rows) is sown with the variety A and the other half with the variety B. On turning the drill at the end, the next strip is sown so that two half-strips of the same variety B are next each other, but care is taken to leave an interval between the two drill strips exactly equal to the gap in the middle of each drill strip between the two varieties. It requires careful steering but it can be done.

When the experimental field is sown, we get first a single half-drill strip of variety A, then two of the other variety B, then two of A and so forth, ending with a half-drill strip of A. This ending is necessary in order to discount any fertility slope from one end to the other of the field. The situation is illustrated in Fig. 1.

Four consecutive half-drill strips form a sandwich. The idea of this arrangement is based on the empirical fact that the changes in soil fertility that are met with in uniform fields which might be chosen for trials are frequently "monotonic", that is the fertility increases gradually, though perhaps not always uniformly, from one end of the field to the other. If this is so, then the variety B may be favoured by its position in the first pair of half-drill strips, but then it will have a disadvantage, of about the same importance, in the second, and so on. Consequently sandwiches, considered as units of the experiment, will be well balanced and are likely to provide equal conditions of comparison of the two varieties.

"Student" considered three different ways of treating statistically the results of half-drill strip experiments.

(*a*) *Method of pairs.* This consists in considering the difference $A_i - B_i$ between the yields of A and B in each of the pairs of half-drill strips as independent observations. Consequently the variance of the mean, $\overline{A-B}$, of such differences would be estimated by, say,

$$s_1^2 = \frac{\Sigma\,(A_i - B_i - \overline{A-B})^2}{2n\,(2n-1)}, \qquad \text{......(1)}$$

where $2n$ denotes the number of pairs of half-drill strips, while n is the number of sandwiches.

(*b*) *Method of sandwiches.* Here the differences, say,

$$\left.\begin{array}{l} A_1 - B_1 - B_2 + A_2 = \Delta_1 \\ A_3 - B_3 - B_4 + A_4 = \Delta_2 \end{array}\right\}, \qquad \ldots\ldots(2)$$

are considered as single independent observations, and the variance of their mean $\bar{\Delta}$ is estimated by

$$s_2^2 = \frac{\Sigma(\Delta_i - \bar{\Delta})^2}{n(n-1)}. \qquad \ldots\ldots(3)$$

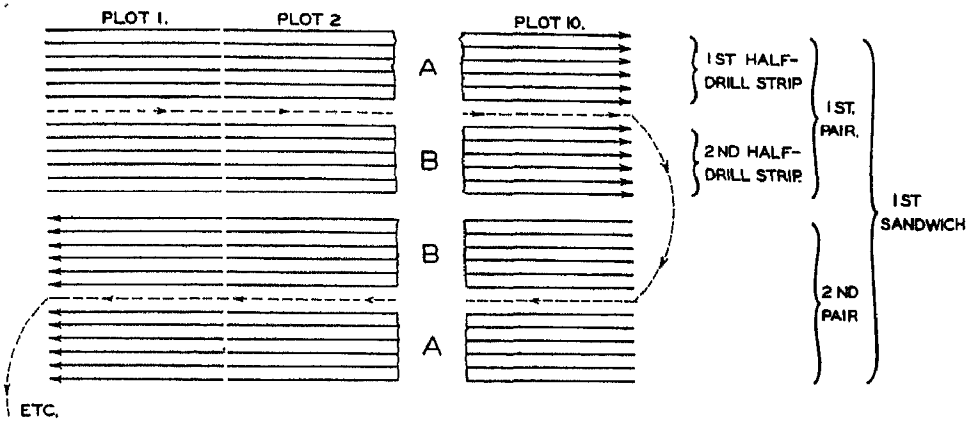


Fig. 1. Directions of sowing half-drill strips with six rows each.

(*c*) Finally, "Student" considered the possibility of artificially multiplying the number of observations by subdividing half-drill strips transversely into several portions. Fig. 1 shows the subdivision of the half-drill strips forming the first sandwich, each into $m = 10$ such portions, which "Student" calls plots. If a_{ij} and b_{ij} denote the yields of the varieties A and B from the jth plots of the ith pair of the half-drill strips, while $\overline{a-b}$ is their arithmetic mean, then its variance could be estimated by

$$s_3^2 = \frac{\Sigma\Sigma\{a_{ij} - b_{ij} - (\overline{a-b})\}^2}{2nm(2nm-1)}. \qquad \ldots\ldots(4)$$

The values such as a_{ij} and b_{ij} are called sheaf weights.

In discussing these three methods, "Student" was aware of the fact that the three formulae for s_1, s_2, and s_3 are not exact, because of the correlations existing between adjacent half-drill strips and between plots adjacent within each of the half-drill strips. Having in view the possibility of linear or nearly linear variation of soil fertility, "Student" indicated that s_1 will tend to overestimate the actual inaccuracy of the experiment. Therefore, he really advised the use of the method of sandwiches. As to the method (*c*), "Student" discussed it at length, concluding that it was likely to be misleading. In fact, the total of the differences $a_{ij} - b_{ij}$ is exactly equal to that of $A_i - B_i$. On the other hand, the number of the former is m times that of the latter. Finally, the cultivation processes and other circumstances result in the fact that the values of $a_{ij} - b_{ij}$, as calculated for the same pair of half-drill strips, tend to be alike. The result is that the greater the number of "plots", the smaller the value of s_3, and this decrease is purely artificial, connected with the method of calculating and not with the accuracy of the experiment. "Student" illustrates these facts on Beaven's data and adds the following footnote which we think is significant:

"A fallacy arising from a similar neglect of correlation has come under my notice in some American work, but there the absurdity is more easily demonstrated. In the *Journal of the American Society of Agronomists*, vol. IX, 1917, p. 138, A. G. McCall proposed that in order to save the trouble of harvesting and weighing 1/10th acre plots, a number of square yards

should be cut out and harvested separately, the square yards being taken systematically throughout the 1/10th acre plot, and the yield per acre calculated from these square yards. So far, so good, by taking enough square yards the slight loss of accuracy may perhaps be made up by gain in time or feasibility of operating. But in 1919, Arny and Steinmetz, *Journal of the American Society of Agronomists*, vol. XI, pp. 88, 89, applying this method, compared the error of the yield calculated from a few square yards cut from each of a number of 1/10th acre plots with that calculated from the 1/10th acre plots themselves. They found it substantially greater, but, say they, by increasing the number of square yards cut from each 1/10th acre plot to n, we can decrease the error in the proportion $1/\sqrt{n}$, and so we can actually determine the yield more accurately by weighing up 10 or 20 square yards than by weighing up the whole half acre. It is rather surprising that they did not realise that there are 484 square yards in 1/10th acre, so that by taking 484 square yards they would be likely to be more accurate than if they took any lesser number and a fortiori tremendously more accurate than they would be if they took the same 484 square yards and called it 1/10th acre! Of course their formula also should be

$$\sigma\sqrt{\frac{1+r(n-1)}{n}}, \qquad \ldots.(5)$$

where r is the correlation between the yields on the square yards composing 1/10th acre plots, and not

$$\sigma\sqrt{\frac{1}{n}}. \qquad \ldots(6)$$

"The same fallacy has been used to extol the 'rod row' method of determining yield, i.e., the method of cutting along the drill a row one rod in length to represent the yield of the plot from which it is cut."

The correlation between sheaf weights (i.e. yields of "plots") within one half-drill strip is clearly seen in Fig. 2, which represents, in the form used by Barbacki and Fisher, the experimental results of Gustav A. Wiebe referred to by "Student" in the present paper.

Wiebe sowed 125 rows of wheat, divided them transversely into twelve portions of 5 yards each, and harvested and weighed separately the yields of the 1500 portions arranged in twelve columns. In order to obtain fuller information on certain points, we wrote to him and he has kindly supplied us with the following details for which we are very grateful: (1) The direction of ploughing was that of the rows, from west to east. (2) The direction followed by the drill, which sowed eight rows at a time, was also *always* from west to east, i.e. it did not proceed backwards and forwards in the usual manner suggested in Fig. 1, but returned outside the field each time to start again from the west. (3) As "Student", after examining the original data, has suggested, there was a fault in the drill not realized until after the experiment was completed, in particular holes Nos. 3 and 6 having a higher grain delivery than the others.

Dealing separately with each column, Barbacki and Fisher totalled the yields of single rows in groups of six, omitting one row between each consecutive pair of groups. The four-figure numbers in the above diagram represent the results they have obtained and used for further calculations. Each number corresponds exactly to what "Student" called a sheaf weight. To test the relative accuracy of systematic half-drill strips, the authors assigned to particular rows of plots imaginary treatments, as shown in Fig. 2. Then they proceeded to calculate (1) the actual difference between the mean of yields of A and B and (2) the estimate of the S.D. according to formula (4), that is to say they followed exactly the method which "Student" was at pains to advise not to follow, as it leads to an underestimation of the S.D. The results obtained by Barbacki and Fisher confirm what "Student" expected: the estimate of error, as calculated from sheaf weights, is too low. It is also easy to see that "Student" was right in giving the reason for the underestimation of the error variance, if sheaf weights are used for this purpose. He pointed out that the plots within the same half-drill strip, i.e. in the same row in Fig. 2, tend to be correlated. They actually are highly correlated in the experimental data used by Barbacki and Fisher. To illustrate this

North

West / East

A 4410 **126**	A 4035 **115**	A 3865 **110**	A 3640 **104**	A 3650 **104**	A 3985 **114**	A 3490 **100**	A 3330 **95**	A 3358 **96**	A 3712 **106**	A 3487 **99**	A 3781 **108**
B 3950 **113**	B 3865 **110**	B 3295 **94**	B 2960 **84**	B 2925 **83**	B 3685 **105**	B 3400 **97**	B 3040 **87**	B 2889 **82**	B 3195 **91**	B 3496 **100**	B 3576 **102**
B 4185 **119**	B 4075 **116**	B 3325 **95**	B 2860 **82**	B 2965 **85**	B 3770 **108**	B 3240 **92**	B 2735 **78**	B 2764 **79**	B 3460 **99**	B 3273 **93**	B 3442 **98**
A 3785 **108**	A 3515 **100**	A 3255 **93**	A 2815 **80**	A 2630 **75**	A 3295 **94**	A 2875 **82**	A 2630 **75**	A 2775 **79**	A 3040 **87**	A 2940 **84**	A 3152 **90**
A 3870 **110**	A 3780 **108**	A 3660 **104**	A 2980 **85**	A 2650 **76**	A 3250 **93**	A 2925 **83**	A 2915 **83**	A 2933 **84**	A 3277 **93**	A 3042 **87**	A 3363 **96**
B 3910 **112**	B 3690 **105**	B 3705 **106**	B 3050 **87**	B 2910 **83**	B 3630 **104**	B 2985 **85**	B 3130 **89**	B 2986 **85**	B 3040 **87**	B 2778 **79**	B 3123 **89**
B 3890 **111**	B 3695 **105**	B 3720 **106**	B 2990 **85**	B 2970 **85**	B 3315 **95**	B 2910 **83**	B 2985 **85**	B 2851 **81**	B 2635 **75**	B 2906 **83**	B 3081 **88**
A 4190 **120**	A 3970 **113**	A 4335 **124**	A 3350 **96**	A 3325 **95**	A 3870 **110**	A 3120 **89**	A 3015 **86**	A 3097 **88**	A 2909 **83**	A 2936 **84**	A 3628 **103**
A 4170 **119**	A 4070 **116**	A 4455 **127**	A 3610 **103**	A 3365 **96**	A 3460 **99**	A 2970 **85**	A 2855 **81**	A 2877 **82**	A 2834 **81**	A 3020 **86**	A 3632 **104**
B 4015 **115**	B 4480 **128**	B 4730 **135**	B 3805 **109**	B 3375 **96**	B 3545 **101**	B 3080 **88**	B 2810 **80**	B 2794 **80**	B 2974 **85**	B 2770 **79**	B 3805 **109**
B 4150 **118**	B 4755 **136**	B 5065 **144**	B 4125 **118**	B 3550 **101**	B 3740 **107**	B 3425 **98**	B 2690 **77**	B 2789 **80**	B 2810 **80**	B 2895 **83**	B 3695 **105**
A 4190 **120**	A 4740 **135**	A 5265 **150**	A 4415 **126**	A 3675 **105**	A 3965 **113**	A 3685 **105**	A 3030 **86**	A 2782 **79**	A 2904 **83**	A 3080 **88**	A 3798* **108**
A 4095 **117**	A 5075 **145**	A 5495 **157**	A 4270 **122**	A 3760 **107**	A 4010 **114**	A 3695 **105**	A 3255 **93**	A 2759 **79**	A 3118 **89**	A 3287 **94**	A 3547 **101**
B 3805 **109**	B 4360 **124**	B 4415 **126**	B 3870 **110**	B 3585 **102**	B 3785 **108**	B 4025 **115**	B 3300 **94**	B 3199 **91**	B 3407 **97**	B 3473 **99**	B 3572 **102**
B 4005 **114**	B 4225 **120**	B 3840 **110**	B 3800 **108**	B 3780 **108**	B 3780 **108**	B 4025 **115**	B 3710 **106**	B 3564 **102**	B 3616 **103**	B 3539 **101**	B 3853 **110**
A 3700 **106**	A 4325 **123**	A 3550 **101**	A 3455 **99**	A 3540 **101**	A 3660 **104**	A 3980 **114**	A 3705 **106**	A 3577 **102**	A 3759 **107**	A 3558 **101**	A 3673 **105**

South

* Probably through a printer's error, this number was given as 2798 in Barbacki and Fisher's Table I; the modified number leads to the total yield which they give in the text.

Fig. 2. Yields of plots on Wiebe's field as arranged by Barbacki and Fisher. N B. The main figures are yields; the figures below in heavier type express each yield as a percentage of the mean.

point, their original figures were calculated as percentages of the average sheaf weight and the results are given in heavier type in the figure. It is seen that, with a single exception, the yields of all plots in the first row exceed those of corresponding plots in the second and that a similar tendency appears in the other rows. Roughly speaking, each column of plots tells about the same story as any other column and as would the sums of all twelve plots along the rows. This is just the circumstance "Student" had in mind when stating that the use of sheaf weights for the estimation of error variance must lead to its underestimation.

Barbacki and Fisher tried to show that the systematically analysed half-drill strips vitiate "Student's" t-test. For this purpose they split their data into six hypothetical experiments. The first of these was supposed to include the first and the seventh columns of plots, the second experiment, the two next columns, second and eighth, etc. They have obtained exactly what could be expected from "Student's" warning concerning the correlation between the sheaf weights within one half-drill strip: each of their hypothetical experiments was an approximate replicate of any other among them, in particular all the six t's were of the same sign.

The application of statistical analysis to Wiebe's data has emphasised two points of importance. In the first place before carrying out a half-drill strip experiment it is very necessary to examine the delivery of the separate holes in the drill. Secondly in using available uniformity trial data to investigate the relative efficiency of various experimental lay-outs, care must be taken that the plots selected are so chosen that possible faults in the drill (often hard to detect after the event) do not add spurious fluctuations in yield to those genuinely due to change in the level of fertility. Had the sheaf weights chosen by Barbacki and Fisher been seven drills broad instead of six, with one intervening row omitted, the correlation effect probably would have been reduced.

It should be mentioned that in a later paper* "Student" has advanced a method of dealing with pairs of half-drill strips, allowing for a linear fertility gradient, which has a definite justification on the grounds of the theory of probability. In fact it is a direct consequence of the following system of hypotheses:

(*a*) Within each pair of half-drill strips, the fertility of soil varies linearly and the slope is the same for all pairs.

(*b*) The treatments compared react similarly to changes in soil fertility; if a strip, P_1, is better than another, P_2, for one treatment, then it is also to about the same extent better for the other treatment.

(*c*) Technical errors of experimentation are independent of changes in soil fertility and of the treatments, and are normally distributed about zero.

It is, of course, uncertain whether in any particular case the changes in the fertility level over the field can be represented with sufficient accuracy by portions of straight lines with constant slope. But this unescapable difficulty is of a kind which arises always whenever we try to deal mathematically with any objects of the outside world. Mathematics deals with mathematical conceptions, not with real things and we can expect no more than a certain amount of correspondence between the two. This applies equally to mathematical treatment of other experimental designs such as randomized blocks and Latin squares in spite of the random assignment of treatments, though strong opinions have sometimes been expressed to the contrary. Whether in the case of the half-drill strips, this correspondence is, for practical purposes, satisfactory or not must be tested empirically. This kind of test, having in view the question as to whether the above hypotheses (*a*), (*b*) and (*c*) do usually lead to a satisfactory approximation of the actual level of fertility on experimental fields, is now being carried out and it is hoped that the results will soon be published.

2. *The balanced arrangement of "Student's" section* 3. In this section "Student" has suggested a form of completely balanced arrangement which might be applied to the 192

* This method seems first to have been described and illustrated by "Student" in his article on "Yield Trials" published in *Baillière's Encylopedia of Scientific Agriculture*, II, London (1931), 1358–60. In the algebraic table given later in his paper in the Supplement to *J. roy. Statist. Soc.* III (1936), 121, it seems necessary to read $S(A-B)^2 - n(\overline{A-B})^2$ for $S(A-B)^2$.

plots into which Fisher and Barbacki have chosen to divide Wiebe's field. His main purpose here was to show that for such an arrangement the total treatment difference (513 g.) is comfortably within the standard error (2259). The much larger total treatment difference (5875 g.) given by Fisher and Barbacki (*loc. cit.* p. 191) applies to an arrangement which was not balanced horizontally, and is therefore irrelevant to any fair comparison of balanced versus random arrangements of sheaf weights. In examining the method which "Student" followed in analysing his results, we have not however been able to reproduce exactly the analysis of variance table which he gives on p. 371 above. While the difference between the estimate of error, which he gives (2259) and that which we believe is correct on the hypothesis from which he appears to start (2385), is not in any sense vital to his argument, we think it may be worth while examining the hypothesis regarding the fertility level in the experimental field which seems to underly the statistical treatment of the data that he has outlined.

Consider then the arrangement of $16st$ plots in $4s$ rows and $4t$ columns and assume that two varieties only, A and B, are sown on these plots according to the scheme shown below. For the data of Fig. 2 $s=4$, $t=3$.

	1	2	3	4	...	$4t-1$	$4t$
1	A	B	B	A		B	A
2	B	A	A	B		A	B
3	B	A	A	B		A	B
4	A	B	B	A		B	A
5	A	B	B	A		B	A
⋮							
$4s-1$	B	A	A	B		A	B
$4s$	A	B	B	A		B	A

It will be assumed that the plots are oblong and narrow, their longer side being parallel to the rows of the diagram. Each pair of the plots, with their longer sides adjacent, will provide a comparison between A and B. There will be $2s$ such pairs of plots within each of $4t$ columns, or $8st$ pairs in all. It is clear, however, that any such comparison within a pair of plots by itself would not be a fair one, as the two plots forming the pair are not likely to be of the same fertility. It will be explicitly assumed that their fertility is different and we shall denote by F_{ij} for $i=1, 2, \ldots, 2s$, $j=1, 2, \ldots, 4t$, the advantage of the lower plot over the upper within any of $8st$ pairs into which the whole field may be divided. We shall further denote by $F_{..}$, the average of all the numbers F_{ij}, by $F_{i.}$ their average within the ith row and by $F_{.j}$ the average within the jth column. Consequently, if we take into consideration the ith pair of plots within the jth column (pair (ij) for short) and denote by A_{ij} the average of the true yields which the variety A is able to give on these plots, then its true yield on the upper plot would be

$$A_{ij}-\tfrac{1}{2}F_{ij} \qquad \ldots\ldots(7)$$

and that on the lower one

$$A_{ij}+\tfrac{1}{2}F_{ij}. \qquad \ldots\ldots(8)$$

This could be written for any variety and for any field; the above formulae are meant to explain the notation, but do not imply any hypothesis concerning the experimental field or varieties. We shall now formulate the hypotheses leading to the new method suggested by "Student", which we assume he must have had in mind when writing his paper.

(1) We shall assume that the advantage of the lower over the upper plot in a pair is of the same magnitude for both varieties, A and B. Consequently, the true yields of the variety B on the plots of the (ij) pair will be

$$A_{ij}-\tfrac{1}{2}F_{ij}-\Delta \quad\text{and}\quad A_{ij}+\tfrac{1}{2}F_{ij}-\Delta, \qquad \ldots\ldots(9)$$

respectively, where Δ denotes the difference between the yields $A-B$ if sown in identical conditions.

(2) We shall make a certain hypothesis regarding the variability of soil over the field or,

what comes to the same thing, regarding the values of the F_{ij}. Notice first that whatever the field may be, we can write

$$\begin{aligned} F_{ij} &= F_{..} + (F_{i.} - F_{..}) + (F_{.j} - F_{..}) + (F_{ij} - F_{i.} - F_{.j} + F_{..}) \\ &= F_{..} + R_i + C_j + \eta_{ij}, \end{aligned} \qquad \ldots\ldots(10)$$

where R_i means the correction to be added to $F_{..}$ to obtain the average of the F_{ij} within the ith row of pairs and C_j a similar correction for columns. We shall assume that the selection of the field for the experiment was a careful one so that the variation of fertility over it is regular, in the sense that the first three terms in the right-hand side of (10) represent sufficiently accurately the left-hand side, thus

$$F_{ij} = F_{..} + R_i + C_j \quad \text{for } i=1, 2, \ldots, 2s; j=1, 2, \ldots, 4t. \qquad \ldots\ldots(11)$$

It will be noticed that this implies

$$\Sigma(R_i) = \Sigma(C_j) = 0. \qquad \ldots\ldots(12)$$

Substituting (11) into (7), (8) and (9), we shall obtain the hypothetical set-up of "true yields" of the varieties A and B, which they are able to give on both plots of the (ij) pair

$$\left.\begin{aligned} &A_{ij} - \tfrac{1}{2}(F_{..} + R_i + C_j), \quad A_{ij} - \tfrac{1}{2}(F_{..} + R_i + C_j) - \Delta, \\ &A_{ij} + \tfrac{1}{2}(F_{..} + R_i + C_j), \quad A_{ij} + \tfrac{1}{2}(F_{..} + R_i + C_j) - \Delta. \end{aligned}\right\} \qquad \ldots\ldots(13)$$

The following table represents the set-up within the pairs of the first few rows and columns.

TABLE I

Expected yield of A and B

	1st column of pairs	2nd column of pairs	3rd column of pairs	4th column of pairs
1st row of pairs	$A_{11} - \frac{1}{2}(F_{..} + R_1 + C_1)$ $A_{11} + \frac{1}{2}(F_{..} + R_1 + C_1) - \Delta$	$A_{12} - \frac{1}{2}(F_{..} + R_1 + C_2) - \Delta$ $A_{12} + \frac{1}{2}(F_{..} + R_1 + C_2)$	$A_{13} - \frac{1}{2}(F_{..} + R_1 + C_3) - \Delta$ $A_{13} + \frac{1}{2}(F_{..} + R_1 + C_3)$	$A_{14} - \frac{1}{2}(F_{..} + R_1 + C_4)$ $A_{14} + \frac{1}{2}(F_{..} + R_1 + C_4) - \Delta$
2nd row of pairs	$A_{21} - \frac{1}{2}(F_{..} + R_2 + C_1) - \Delta$ $A_{21} + \frac{1}{2}(F_{..} + R_2 + C_1)$	$A_{22} - \frac{1}{2}(F_{..} + R_2 + C_2)$ $A_{22} + \frac{1}{2}(F_{..} + R_2 + C_2) - \Delta$	$A_{23} - \frac{1}{2}(F_{..} + R_2 + C_3)$ $A_{23} + \frac{1}{2}(F_{..} + R_2 + C_3) - \Delta$	$A_{24} - \frac{1}{2}(F_{..} + R_2 + C_4) - \Delta$ $A_{24} + \frac{1}{2}(F_{..} + R_2 + C_4)$
3rd row of pairs	$A_{31} - \frac{1}{2}(F_{..} + R_3 + C_1)$ $A_{31} + \frac{1}{2}(F_{..} + R_3 + C_1) - \Delta$	$A_{32} - \frac{1}{2}(F_{..} + R_3 + C_2) - \Delta$ $A_{32} + \frac{1}{2}(F_{..} + R_3 + C_2)$	$A_{33} - \frac{1}{2}(F_{..} + R_3 + C_3) - \Delta$ $A_{33} + \frac{1}{2}(F_{..} + R_3 + C_3)$	etc.
	etc.	etc.		

(3) The third hypothesis concerns the difference, ϵ, between these "true yields" and those which might be observed. It will be assumed that the formulae (13) make a full allowance for both differences between the varieties (Δ) and the soil variation (components A_{ij}, $F_{..}$, R's and C's) and that therefore the difference ϵ is due only to inevitable random technical errors of experimentation, normally distributed about zero with an unknown variance.

Are these hypotheses satisfied in practice? It is difficult to say for certain but it seems likely that they might be. In fact, the number of arbitrary constants is very considerable and the range of variety in fertility levels which could be constructed by varying them, even within narrow limits, is enormous. Nor has the scheme any sort of rigidity involved in the original "Student's" set up for the half-drill strips, where the fertility slope was assumed to be constant throughout the field. Now the slope can change from pair to pair within each column and row, being sometimes negative and sometimes positive. All these considerations

suggest that if the choice of the experimental field is not very unlucky, then the changes of its fertility can be approximated by the above scheme with a great degree of accuracy.

Granting this, we may proceed to the analysis of the data in Fig. 2 as follows: write y_{ij} for the difference (observed yield in lower plot) – (observed yield in upper plot) of the (i, j) pair. It will be noted that the $8st$ values of y_{ij} will fall into two sets:

Set (α), say, containing the $4st$ values which have been obtained by subtracting the yield of variety B from that of variety A. This will include from Table II y_{12}, y_{13}, y_{21} and y_{24}.

Set (β), containing the remaining values, for which the yield of A has been subtracted from that of B; e.g. y_{11}, y_{14}, y_{22} and y_{23} in Table II.

The $8st$ differences of observed yields may be set out in the following scheme:

TABLE II

	1st column of pairs	2nd column of pairs	3rd column of pairs	4th column of pairs		Expectation of total
1st row of pairs	$y_{11}=$ $F_{..}+R_1+C_1-\Delta+u_{11}$	$y_{12}=$ $F_{..}+R_1+C_2+\Delta+u_{12}$	$y_{13}=$ $F_{..}+R_1+C_3+\Delta+u_{13}$	$y_{14}=$ $F_{..}+R_1+C_4-\Delta+u_{14}$		$4t(F_{..}+R_1)$
2nd row of pairs	$y_{21}=$ $F_{..}+R_2+C_1+\Delta+u_{21}$	$y_{22}=$ $F_{..}+R_2+C_2-\Delta+u_{22}$	$y_{23}=$ $F_{..}+R_2+C_3-\Delta+u_{23}$	$y_{24}=$ $F_{..}+R_2+C_4+\Delta+u_{24}$	etc.	$4t(F_{..}+R_2)$
			etc.			
Expectation of totals	$2s(F_{..}+C_1)$	$2s(F_{..}+C_2)$	$2s(F_{..}+C_3)$	$2s(F_{..}+C_4)$		$8st F_{..}$

Here the u_{ij} are the differences of the ϵ's referred to above, and are supposed normally and randomly distributed about zero with unknown variance σ^2. We shall now write $y_{..}$ for the grand mean of the y's, $y_{i.}$ for the mean y in the ith row, $y_{.j}$ for the mean y in the jth column, $y'_{..}$ for the mean of the $4st$ y's of set (α) and $y''_{..}$ for the mean of the $4st$ y's of set (β). Notice that

$$y_{..}=\tfrac{1}{2}(y'_{..}+y''_{..}),\quad \tfrac{1}{2}(y'_{..}-y''_{..})=y_{..}-y''_{..}=-(y_{..}-y'_{..}).$$

With these definitions it may be shown by application either of the Markoff theorem or the usual procedure for testing linear hypotheses, that

(a) $y_{..}$ is an unbiased estimate of $F_{..}$

(b) $y_{i.}-y_{..}$ is an unbiased estimate of R_i

(c) $y_{.j}-y_{..}$ is an unbiased estimate of C_j

(d) $\tfrac{1}{2}(y'_{..}-y''_{..})$ is an unbiased estimate of Δ.

(e) The differences, for set (α),

$$y_{ij}-y_{..}-(y_{i.}-y_{..})-(y_{.j}-y_{..})-\tfrac{1}{2}(y'_{..}-y''_{..})=y_{ij}-y_{i.}-y_{.j}+y''_{..},$$

for set (β) $$y_{ij}-y_{..}-(y_{i.}-y_{..})-(y_{.j}-y_{..})+\tfrac{1}{2}(y'_{..}-y''_{..})=y_{ij}-y_{i.}-y_{.j}+y'_{..},$$

are normally distributed about zero, and if we write the sums of squares

$$S_0^2=\sum_i\sum_j^{\alpha}(y_{ij}-y_{i.}-y_{.j}+y''_{..})^2+\sum_i\sum_j^{\beta}(y_{ij}-y_{i.}-y_{.j}+y'_{..})^2, \qquad \text{.....(14)}$$

then the expected value, $$E(S_0^2)=(8st-2s-4t)\sigma^2. \qquad \text{......(15)}$$

Hence we obtain the following partition of the total sum of squares.

$$\sum_j\sum_i(y_{ij}^2)=8sty_{..}^2+4t\sum_i(y_{i.}-y_{..})^2+2s\sum_j(y_{.j}-y_{..})^2+8st\left(\frac{y_{..}-y''_{..}}{2}\right)^2+S_0^2. \qquad \text{..... (16)}$$

Degrees of freedom:

$$8st=\ 1\ +\ (2s-1)\ +\ (4t-1)\ +\ 1\ +(8st-2s-4t)$$

Applying this theory to the 96 differences obtained by subtracting the upper plot from the lower plot value for all the pairs of Fig. 2, and remembering that $s = 4$, $t = 3$, we reach the following analysis of variance table:

TABLE III

Sum of squares due to		Degrees of freedom	Mean square
Average fertility slope for whole field	157221	1	
Vertical (row) fertility slopes	4351285	7	
Horizontal (column) fertility slopes	373246	11	
Varietal difference	2741	1	
Residual	4502606	76	59245
Total	9387099	96	

The estimate of σ obtained from the residual is therefore 243·4. The total treatment difference is $48(y'_{..} - y''_{..}) = 513$ g., which has an estimated standard error of

$$\sqrt{(96)} \times 243{\cdot}4 = 2385 \text{ g.}$$

"Student's" point that the difference is well within the standard error is therefore established.

On comparing Table III with that given by "Student" on p. 371 above, it will be seen that the sum of our 1st and 2nd row is equal to his 2nd row, while our 4th row agrees with his 3rd row. The method he has used for extracting the horizontal fertility slopes was however probably not justifiable, and consequently his residual sum of squares is too small, as also his estimate (2259) of the standard error of the total treatment difference. In stating on p. 371 and also in the introductory remarks on p. 363 that a balanced arrangement gives "a slightly smaller error" than the randomised one he appears to have been at fault; these points, however, he would have undoubtedly cleared up before the paper was printed.

We may conclude this note by considering the consequences of a disagreement between the theoretical set-up concerning the level of fertility and what happens in practice. If the theoretical set-up is ideally correct, then any of the expressions

$$y_{ij} - y_{i.} - y_{.j} + y''_{..} \quad \text{or} \quad y_{ij} - y_{i.} - y_{.j} + y'_{..}, \qquad \ldots(17)$$

will vary about zero with their standard deviations equal to $\sigma\sqrt{\{(8st - 2s - 4t)/8st\}}$. If, however, the theoretical model fails to represent the actual fertility changes in the field, then some at least of the expressions (17) will vary about means m_{ij}, say, different from zero. Consequently the expected value of the S_0^2 of (14) could be broken up into two components, one of which will be the estimate of σ^2 multiplied by the degrees of freedom, and the other, always positive, will be proportional to $\Sigma\Sigma(m_{ij}^2)$. Thus if the model of the soil fertility is not correct, which strictly speaking will always be the case, the expression for S_0^2 will tend to over-estimate the actual error variance. This effect seems to have been what "Student" had in mind. How large or how frequent such over-estimations may be it is impossible to say, except by special inquiry on numerous uniformity trial data.

Similar considerations apply to the effect of the discrepancy between the mathematical model and the actual changes in fertility level on the estimate of the mean difference Δ. But here we may notice that whatever could be said against the chances of errors due to changes in soil fertility cancelling out in the estimate $\frac{1}{2}(y'_{.} - y''_{.})$ calculated from "Student's" suggested lay-out, could be applied with as much or more emphasis to many other experimental arrangements, such as for example the Latin Square.

There remains, of course, the question raised by "Student" himself as to whether and in what cases this new lay-out would be practical.

THE FIRST SIX MOMENTS OF χ^2 FOR AN n-FOLD TABLE WITH n DEGREES OF FREEDOM WHEN SOME EXPECTATIONS ARE SMALL

By J. B. S. HALDANE

HALDANE (1937) gave the first four moments of this distribution. They were derived as a special case of a more general formula. They may, however, be derived directly, by a simple process which allows of the calculation of higher moments with relative ease.

Consider a large sample, in which the number of individuals of a certain type, or the number of expected successes if all the experiments in the sample are independent, is m, where m is small compared with the number in the sample. Then if x be the observed number of individuals of the type considered, or the number of successes in the sample, the probability of observing x is $\frac{m^x}{e^m x!}$, the probabilities forming a Poisson series. It can further readily be shown that the moment-generating function of such a series, moments being taken about zero, is $e^{m(e^t-1)}$. If moments are taken about the mean m, this function is $e^{m(e^t-1-t)}$. The cumulant-generating function is therefore $m(e^t-1)$, and all cumulants are equal to m. The moments about the mean are most readily calculated from the cumulants by equations which have been given by Fisher (1928) up to μ_8, and are continued below up to μ_{12}. These equations, which hold for any distribution, are as follows for the even moments, which alone concern us:

$$\mu_2=\kappa_2,$$

$$\mu_4=\kappa_4+3\kappa_2^2,$$

$$\mu_6=\kappa_6+5(3\kappa_4\kappa_2+2\kappa_3^2)+15\kappa_2^3,$$

$$\mu_8=\kappa_8+7(4\kappa_6\kappa_2+8\kappa_5\kappa_3+5\kappa_4^2)+70(3\kappa_4\kappa_2^2+4\kappa_3^2\kappa_2)+105\kappa_2^4,$$

$$\begin{aligned}\mu_{10}=\kappa_{10}&+3(15\kappa_8\kappa_2+40\kappa_7\kappa_3+70\kappa_6\kappa_4+42\kappa_5^2)\\&+105(6\kappa_6\kappa_2^2+24\kappa_5\kappa_3\kappa_2+15\kappa_4^2\kappa_2+20\kappa_4\kappa_3^2)\\&+3150(\kappa_4\kappa_2^3+2\kappa_3^2\kappa_2^2)+945\kappa_2^5,\end{aligned}$$

$$\begin{aligned}\mu_{12}=\kappa_{12}&+11(6\kappa_{10}\kappa_2+20\kappa_9\kappa_3+45\kappa_8\kappa_4+72\kappa_7\kappa_5+42\kappa_6^2)\\&+33(45\kappa_8\kappa_2^2+240\kappa_7\kappa_3\kappa_2+420\kappa_6\kappa_4\kappa_2+280\kappa_6\kappa_3^2+252\kappa_5^2\kappa_2\\&\qquad+840\kappa_5\kappa_4\kappa_3+175\kappa_4^3)\\&+385(36\kappa_6\kappa_2^3+216\kappa_5\kappa_3\kappa_2^2+135\kappa_4^2\kappa_2^2+360\kappa_4\kappa_3^2\kappa_2+40\kappa_3^4)\\&+17{,}325(3\kappa_4\kappa_2^4+8\kappa_3^2\kappa_2^3)+10{,}395\,\kappa_2^6.\end{aligned}$$

In these equations the coefficient of $\kappa_a^\alpha \kappa_b^\beta \kappa_c^\gamma \ldots$ is

$$\frac{(a\alpha+b\beta+c\gamma+\ldots)!}{\alpha!\,\beta!\,\gamma!\ldots(a!)^\alpha(b!)^\beta(c!)^\gamma\ldots}.$$

Terms bracketed together are multiples of the same power of m in the case of the Poisson distribution and therefore

$$\begin{aligned}
\mu_2 &= m,\\
\mu_4 &= m+3m^2,\\
\mu_6 &= m+25m^2+15m^3,\\
\mu_8 &= m+119m^2+490m^3+105m^4,\\
\mu_{10} &= m+501m^2+6{,}825m^3+9{,}450m^4+945m^5,\\
\mu_{12} &= m+2{,}035m^2+74{,}316m^3+302{,}995m^4+190{,}575m^5+10{,}395m^6.
\end{aligned}$$

These are the even moments of $(x-m)$ about its mean, zero. They are therefore the successive moments of $(x-m)^2$ about zero, which is not its mean. The means of powers, or moments about zero, of $\chi^2=\dfrac{(x-m)^2}{m}$, for a single degree of freedom, are therefore

$$\begin{aligned}
\mu_1' &= 1,\\
\mu_2' &= 3+m^{-1},\\
\mu_3' &= 15+25m^{-1}+m^{-2},\\
\mu_4' &= 105+490m^{-1}+119m^{-2}+m^{-3},\\
\mu_5' &= 945+9{,}450m^{-1}+6{,}825m^{-2}+501m^{-3}+m^{-4},\\
\mu_6' &= 10{,}395+190{,}575m^{-1}+302{,}995m^{-2}+74{,}316m^{-3}+2{,}035m^{-4}+m^{-5}.
\end{aligned}$$

The moments about the mean, 1, are

$$\begin{aligned}
\mu_2 &= 2+m^{-1},\\
\mu_3 &= 8+22m^{-1}+m^{-2},\\
\mu_4 &= 60+396m^{-1}+115m^{-2}+m^{-3},\\
\mu_5 &= 544+7{,}240m^{-1}+6{,}240m^{-2}+496m^{-3}+m^{-4},\\
\mu_6 &= 6{,}040+140{,}740m^{-1}+263{,}810m^{-2}+71{,}325m^{-3}+2{,}029m^{-4}+m^{-5}.
\end{aligned}$$

The cumulants are

$$\begin{aligned}
\kappa_1 &= 1,\\
\kappa_2 &= 2+m^{-1},\\
\kappa_3 &= 8+22m^{-1}+m^{-2},\\
\kappa_4 &= 48+384m^{-1}+112m^{-2}+m^{-3},\\
\kappa_5 &= 384+6{,}720m^{-1}+6{,}000m^{-2}+486m^{-3}+m^{-4},\\
\kappa_6 &= 3{,}840+21{,}300m^{-1}+249{,}600m^{-2}+69{,}160m^{-3}+2{,}004m^{-4}+m^{-5}.
\end{aligned}$$

Hence if we have a series of n samples, in the rth of which the expectation is m_r, and if $R_s=\sum_{=1}^{n} m_r^{-s}$, then the cumulants of $\chi^2=\sum_{=1}^{n}(x_r-m_r)^2/m_r$ are

$$\left.\begin{aligned}
\kappa_1 &= n, \\
\kappa_2 &= 2n + R_1, \\
\kappa_3 &= 8n + 22R_1 + R_2, \\
\kappa_4 &= 48n + 384R_1 + 112R_2 + R_3, \\
\kappa_5 &= 384n + 6720R_1 + 6{,}000R_2 + 486R_3 + R_4, \\
\kappa_6 &= 3{,}840n + 21{,}300R_1 + 249{,}600R_2 + 69{,}160R_3 + 2{,}004R_4 + R_5.
\end{aligned}\right\} \quad \ldots\ldots(1)$$

The successive moments are therefore

$$\begin{aligned}
\mu_2 &= 2n + R_1, \\
\mu_3 &= 8n + 22R_1 + R_2, \\
\mu_4 &= 12n(n+4) + 12(n+32)R_1 + R_1^2 + 112R_2 + R_3, \\
\mu_5 &= \kappa_5 + 10\kappa_3\kappa_2, \\
\mu_6 &= \kappa_6 + 15\kappa_4\kappa_2 + 10\kappa_3^2 + 15\kappa_2^3.
\end{aligned}$$

The higher moments are very considerably increased, even when m is as large as 5. In this case

μ_2 is increased from		$2n$ to $2{\cdot}2n$,
μ_3	,,	$8n$ to $12{\cdot}44n$,
μ_4	,,	$12n^2 + 48n$ to $14{\cdot}44n^2 + 129{\cdot}288n$,
μ_5	,,	$160n^2 + 384n$ to $273{\cdot}68n^2 + 1{,}971{\cdot}89n$,
μ_6	,,	$120n^3 + 2{,}080n^2 + 3{,}840n$ to $159{\cdot}72n^3 + 5{,}814{\cdot}04n^2 + 18{,}640{\cdot}49n$.

It will, however, be noticed that when n is large the corrections are relatively small, since the leading term is a multiple of $\kappa_2^{\frac{1}{2}r}$ or $\kappa_2^{\frac{1}{2}(r-1)}\kappa_3$, and the first two cumulants are less affected than the later ones. The coefficients in equations (1) occur in the expressions for the cumulants of χ^2 in other cases, and may therefore be used as checks on any calculations of them.

It will be noticed that the new distribution deviates further from normality than the χ^2 distribution for large expectations. If $\gamma_1 = (\beta_1)^{\frac{1}{2}} = \mu_3\mu_2^{-\frac{3}{2}}$ and

$$\gamma_2 = \beta_2 - 3 = \mu_4\mu_2^{-2} - 3$$

are the measures of deviation when expectations are large, and γ_1', γ_2' the same when expectations are small, then $\dfrac{\gamma_1'}{\gamma_1}$ approximates to $1 + 2R_1n^{-1}$, and $\dfrac{\gamma_2'}{\gamma_2}$ approximates to $1 + 7R_1n^{-1}$.

REFERENCES

FISHER, R. A. (1928). "Moments and product-moments of sampling distributions." *Proc. Lond. Math. Soc.* XXX, 199–238.

HALDANE, J. B. S. (1937). "The exact value of the moments of the distribution of χ^2, used as a test of goodness of fit, when expectations are small." *Biometrika*, XXIX, 133–43.

THE APPROXIMATE NORMALIZATION OF A CLASS OF FREQUENCY DISTRIBUTIONS

By J. B. S. HALDANE

One of the central problems of practical statistics is to determine the probability that the divergence of a variable from its expected value should be due to sampling error. For this purpose we require, in general, a table of the integral, or of the sum between certain limits, of the distribution function of the variable in question.

Such tables have been drawn up for the normal distribution and for a certain number of others. When the frequency distribution depends on a single arbitrary parameter, and especially when, as with the χ^2 distribution, this parameter is an integer, tabulation is not very difficult. But where several arbitrary parameters are concerned it is very tedious, and where the number of arbitrary parameters is large, it is quite impracticable. Such a case arises with the χ^2 distribution when expectations are small (Haldane, 1937).

It has, however, long been known that in many cases the distribution of a statistic derived from a sample of n members of a population, or from an experiment repeated n times, tends to normality when n is large. The deviations from normality can sometimes be conveniently expressed by means of Hermitian polynomials. It will be shown that in a large group of cases a simple transformation of the statistic causes its distribution to approximate very much more rapidly than before to normality when n increases.

One of the two transformations here described was first given by Wilson and Hilferty (1931) in the case of χ^2. It will be shown that this transformation may be applied to other distributions as well. And a second transformation, which in some cases is more powerful, will be described. It will be shown that the distribution of the transformed variate often tends so rapidly to normality when n increases that Sheppard's table of the probability integral is entirely sufficient for practical purposes.

Cumulative statistics

An important class of statistics has the following property. If x_m and x_n are the values of the statistic derived from m members of a population (or m experiments) and n different members or experiments, then $x_{m+n}=x_m+x_n$. That is to say x_m is the sum of m independently ascertained values of the variate x. Let $df=f(x)\,dx$ be the distribution of the variate in the population sampled. Where the distribution of x is wholly or partly discontinuous the appropriate expression can readily be given. And let κ_r be the rth cumulant or semi-invariant of this distribution, defined as the coefficient of $\frac{t^r}{r!}$ in the expansion of $\log\int_{-\infty}^{\infty} e^{tx} f(x)\,dx$.

Then if $\kappa_{r,n}$ be the rth cumulant of the distribution of x_n, it can easily be shown that $\kappa_{r,n} = n\kappa_r$.

Statistics with this property may be called cumulative statistics. The best known examples of them are (i) the number of successes in n trials when the probability of success is constant, and (ii) χ^2. They have the property that all their cumulants tend to infinity with n, though the moments after μ_3 tend to infinity with higher powers of n.

Another important class may be called derived cumulative statistics. They have the property that when multiplied by n, or some power of n, all their non-zero cumulants tend to infinity with n. Thus the mean of a sample clearly has the property, and the cumulants of n times the second moment of a sample of n are

$$(n-1)\,\kappa_2; \quad 2\,(n-1)\,\kappa_2^2 + \frac{(n-1)^2}{n}\kappa_4;$$

$$8\,(n-1)\,\kappa_2^3 + \frac{4\,(n-1)\,(n-2)}{n}\kappa_3^2 + \frac{12\,(n-1)^2}{n}\kappa_2\kappa_4 + \frac{(n-1)^3}{n^2}\kappa_6;$$

and so on, all of which tend to infinity with n. When the expected value of a derived cumulative statistic is constant, its rth cumulant tends to zero with n^{1-r}. The transformations here described are applicable to cumulative and derived cumulative statistics provided that their distribution is not symmetrical, or more accurately provided κ_3 does not vanish.

WILSON AND HILFERTY'S TRANSFORMATION OF χ^2

The rth cumulant of the χ^2 distribution for n degrees of freedom is $n\,(r-1)!\,2^{r-1}$. So the first six cumulants are: $\kappa_1 = n$, $\kappa_2 = 2n$, $\kappa_3 = 8n$, $\kappa_4 = 48n$, $\kappa_5 = 384n$, $\kappa_6 = 3840n$, and the corresponding moments about the mean, n, are:

$$\mu_2 = 2n,\ \mu_3 = 8n,\ \mu_4 = 12n\,(n+4),\ \mu_5 = 32n\,(5n+12),\ \mu_6 = 40n\,(3n^2+52n+96).$$

With Wilson and Hilferty we put $\chi^2 = n + x$, and investigate the distribution of $y = \left(\frac{\chi^2}{n}\right)^h$, where h is a constant to be chosen so as to give an approximately normal distribution. We find

$$y = \left(1 + \frac{x}{n}\right)^h$$

$$= 1 + \frac{hx}{n} + h\,(h-1)\frac{x^2}{2!\,n^2} + h\,(h-1)\,(h-2)\frac{x^3}{3!\,n^3} + \dots,$$

which converges provided $|x| < n$.

But the mean value of x^r is μ_r, the rth moment of χ^2 about its mean.

Hence

$$\bar{y}=1+h(h-1)\frac{\mu_2}{2!\,n^2}+h(h-1)(h-2)\frac{\mu_3}{3!\,n^3}+\dots$$

$$=1+h(h-1)\frac{1}{n}+h(h-1)(h-2)\frac{4}{3n^2}+h(h-1)(h-2)(h-3)\frac{(n+4)}{2n^3}$$

$$+h(h-1)(h-2)(h-3)(h-4)\frac{4n(5n+12)}{15n^4}$$

$$+h(h-1)(h-2)(h-3)(h-4)(h-5)\frac{(3n^2+52n+96)}{18n^5}+O(n^{-4})$$

$$=1+h(h-1)n^{-1}+\tfrac{1}{6}h(h-1)(h-2)(3h-1)n^{-2}$$
$$+\tfrac{1}{6}h^2(h-1)^2(h-2)(h-3)n^{-3}+O(n^{-4}) \qquad \dots\dots(1)$$

It may be remarked that this series is always convergent when n is sufficiently large, since μ_r is of order $n^{\frac{1}{2}r}$ when r is even, and $n^{\frac{1}{2}(r-1)}$ when n is odd.

Wilson and Hilferty (1931, p. 686, last line) give an expression which, putting their $p^{-1}=h$, is equivalent to

$$\bar{y}=1+h(h-1)n^{-1}+\tfrac{4}{3}h(h-1)(h-2)n^{-2}+\dots.$$

It would seem that they neglected the contribution of μ_4 to the coefficient of n^{-2}. This does not, however, affect the validity of their transformation (5). Indeed, it is a better approximation than would appear from a first reading of their paper.

By substituting rh for h in equation (1) we can readily find the mean of y^r as

$$\overline{y^r}=1+rh(rh-1)n^{-1}+\tfrac{1}{6}rh(rh-1)(rh-2)(3rh-1)n^{-2}+\dots.$$

Hence we can calculate the moments of y about its mean $\bar{y}$. For example, the third moment is $\mu_3'=4(3h-1)n^{-2}+O(n^{-3})$.

With Wilson and Hilferty we put $h=\frac{1}{3}$, so that the term of order n^{-2} vanishes, and find

$$\begin{aligned}
\bar{y}&=1-2.3^{-2}n^{-1}+80.3^{-7}n^{-3}+O(n^{-4}),\\
\bar{y}^2&=1-2.3^{-2}n^{-1}+4.3^{-4}n^{-2}+56.3^{-7}n^{-3}+O(n^{-4}),\\
\bar{y}^3&=1,\\
\bar{y}^4&=1+4.3^{-2}n^{-1}-4.3^{-3}n^{-2}+80.3^{-7}n^{-3}+O(n^{-4}).
\end{aligned}$$

Hence

$$\left.\begin{aligned}
\text{Mean}=\kappa_1'&=1-2.3^{-2}n^{-1}+80.3^{-7}n^{-3}+O(n^{-4}),\\
\mu_2'=\kappa_2'&=2.3^{-2}n^{-1}-104.3^{-7}n^{-3}+O(n^{-4}),\\
\mu_3'=\kappa_3'&=32.3^{-6}n^{-3}+O(n^{-4}),\\
\mu_4'&=4.3^{-3}n^{-2}-16.3^{-6}n^{-3}+O(n^{-4}),\\
\kappa_4'&=16.3^{-6}n^{-3}+O(n^{-4}).
\end{aligned}\right\} \qquad \dots\dots(2)$$

Here and throughout we shall use accents to denote moments and cumulants of a transformed distribution. The standard deviation is therefore

$$\sigma' = \frac{1}{3}\sqrt{\frac{2}{n}}\left(1 - \frac{26}{243n^2}\right) + O\left(n^{-\frac{7}{2}}\right),$$

and the first two measures of deviation from normality are

$$\begin{aligned} \gamma_1' &= 2^{\frac{7}{2}}\,3^{-3}\,n^{-\frac{3}{2}} + O\left(n^{-\frac{5}{2}}\right) \\ &= 0{\cdot}4190\,n^{-\frac{3}{2}} + O\left(n^{-\frac{5}{2}}\right), \\ \gamma_2' &= -4.3^{-2}\,n^{-1} + O\left(n^{-2}\right) \\ &= -0{\cdot}4444\,n^{-1} + O\left(n^{-2}\right). \end{aligned}$$

These measures are the third and fourth cumulants of the distribution reduced by making its standard deviation unity, and vanish in the case of a normal distribution i.e. $\gamma_1' = \sqrt{\beta_1} = \mu_3'(\mu_2')^{-\frac{3}{2}}$, and $\gamma_2' = \beta_2 - 3 = \kappa_4'(\kappa_2')^{-2}$. They may be compared with the corresponding measures of deviation from normality of the χ^2 distribution itself, namely

$$\begin{aligned} \gamma_1 &= 2^{\frac{3}{2}}\,n^{-\frac{1}{2}} + O\left(n^{-\frac{3}{2}}\right) \\ &= 2{\cdot}828\,n^{-\frac{1}{2}} + O\left(n^{-\frac{3}{2}}\right), \end{aligned}$$

and

$$\gamma_2 = 12\,n^{-1} + O\left(n^{-2}\right).$$

The transformation thus not only reduces γ_1 by a factor $4/(27n)$, but also reduces $|\gamma_2|$ by a factor $1/27$. For this reason it produces a better approximation to normality, even for small values of n, than might have been expected. Similar transformations of other distributions will not necessarily reduce γ_2. Wilson and Hilferty's table shows that, even for $n=2$, the values of χ^2 in the neighbourhood of $P=0{\cdot}05$ and $P=0{\cdot}01$ given by taking the terms of order n^{-1} only for the mean and variance are correct within 1 %. A convenient expression of their result is to say that the variable

$$\xi = \left[\left(\frac{\chi^2}{n}\right)^{\frac{1}{3}} + \frac{2}{9n} - 1\right]\left(\frac{9n}{2}\right)^{\frac{1}{2}} \qquad \text{......(3)}$$

is almost normally distributed with mean zero and standard deviation unity. This variable has been used in genetics by de Winton and Haldane (1935). However since the distribution of χ^2 has already been tabulated for values of n up to 30 it is unlikely that equation (3) will be much used in practice.

Generalization of Wilson and Hilferty's Theorem.

Consider a variable so distributed that, when n tends to infinity, the orders of magnitude of its cumulants are as follows:

$$\kappa_1 = O(n^{1+a}),\ \kappa_2 = O(n^{1+2a}),\ \kappa_3 = O(n^{1+3a}),\ \kappa_4 \leqslant O(n^{1+4a}),$$

and for higher values of r, $\kappa_r \leqslant O(n^{r+ra-4})$, where a is any constant. The departures from normality, if any, are not affected by multiplying the variable by a constant. So multiplying by n^{-a} we obtain a variate so distributed that $\kappa_1 = \kappa_2 = \kappa_3 = O(n)$, $\kappa_4 \leqslant O(n)$, $\kappa_r \leqslant O(n^{r-4})$. In what follows we shall consider the substitution as made.

Our variates include most cumulative and derived cumulative statistics, and some others. Since however cases where $\kappa_3 = 0$ are excluded, we cannot deal with symmetrical distributions, such as "Student's" (1908) distribution of the mean of a sample in terms of its standard deviation, or Fisher's (1928) distribution of the estimate of γ_1. It will also be noted that a number of distributions which tend to normality with n are excluded, for example that of the transformed correlation coefficient.

Let $$x = \kappa_1 + x', \quad \text{and} \quad y = \left(\frac{x}{\kappa_1}\right)^h.$$

Then $$y = \left(1 + \frac{x'}{\kappa_1}\right)^h = 1 + h\frac{x'}{\kappa_1} + h(h-1)\frac{x'^2}{2!\,\kappa_1^2} + \dots.$$

Writing $rh(rh-1)(rh-2)\dots(rh-n+1) = f(r, n)$, we have

$$\overline{y^r} = 1 + f(r,2)\frac{\mu_2}{2!\,\kappa_1^2} + f(r,3)\frac{\mu_3}{3!\,\kappa_1^3} + f(r,4)\frac{\mu_4}{4!\,\kappa_1^4} + \dots,$$

since μ_r, the rth moment of x about its mean, is the mean of x'^r. Hence

$$\overline{y^r} = 1 + f(r,2)\frac{\kappa_2}{2\kappa_1^2} + f(r,3)\frac{\kappa_3}{6\kappa_1^3} + f(r,4)\frac{(3\kappa_2^2 + \kappa_4)}{24\kappa_1^4}$$
$$+ f(r,5)\frac{\kappa_2\kappa_3}{12\kappa_1^5} + f(r,6)\frac{\kappa_2^3}{48\kappa_1^6} + O(n^{-4}). \qquad \dots\dots(4)$$

Since the sum of the suffixes of the cumulants in the numerator is always equal to the power of κ_1 in the denominator it is clear that in this expression factors of the form n^{ra} will cancel out. The term of highest order in which κ_s appears is a multiple of $\kappa_s\kappa_1^{-s}$, which, if $s > 4$ is of order n^{-4} or less, and therefore does not enter into equation (4). The series converges if n is large enough.

It is clear that we can calculate the moments of y about its mean, and choose h, except in one special case, so that the leading term of the third moment vanishes. We can therefore enunciate the following theorem:

"If a variate x be so distributed that its first three cumulants tend to infinity

with n, the fourth with n or more slowly, and in general κ_r with n^{r-4} or more slowly; then if $h = 1 - \frac{\kappa_1 \kappa_3}{3\kappa_2^2}$, $y = \left(\frac{x}{\kappa_1}\right)^h$ is so distributed that its third moment tends to zero with n^{-3}, and hence its γ_1 tends to zero with $n^{-\frac{1}{2}}$."

The statement remains true if x be multiplied by any constant, such as a power of n. But the theorem clearly breaks down if $\kappa_1 \kappa_3 = 3\kappa_2^2$. In this case however the transformation can be performed after adding a suitable constant of order n to the variate. If h has a negative value this may be adjusted by altering the mean as above. Otherwise the approximation (7) given below will clearly break down when x is very small or negative. But this implies that x exceeds its standard deviation $\kappa_2^{\frac{1}{2}}$ by a factor of $|\kappa_1| \kappa_2^{-\frac{1}{2}}$, which is of order $n^{\frac{1}{2}}$. When n is sufficiently large such deviations will be excessively unlikely, and approximation (7) may prove to be valid even when h is negative.

The proof follows. By substituting appropriate values of r in equation (4) we can readily deduce the mean and the first few moments of y. The algebra is tedious but elementary. For example the coefficient of $\frac{3\kappa_2 + \kappa_4}{24\kappa_1^3}$ in the expression for μ_4' is:

$$f(4,4) + 6f(2,4) - 4f(1,4) - 4f(3,4) = 24h^4.$$

In the expression for μ_4' the term of order n^{-4} is given, on the assumption that κ_4 and κ_5 are of order n, and throughout, terms of the same order of magnitude are grouped together. The mean and first few moments about the mean of y are

$$\left.\begin{aligned}
\text{Mean} = \kappa_1' &= 1 + h(h-1)\frac{\kappa_2}{2\kappa_1^2} + h(h-1)(h-2)\frac{[4\kappa_1\kappa_3 + 3(h-3)\kappa_2^2]}{24\kappa_1^4} \\
&\quad + h(h-1)(h-2)(h-3)\frac{[2\kappa_1^2\kappa_4 + 4(h-4)\kappa_1\kappa_2\kappa_3 + (h-4)(h-5)\kappa_2^3]}{48\kappa_1^6} + O(n^{-4}), \\
\mu_2' = \kappa_2' &= \frac{h^2\kappa_2}{\kappa_1^2} + h^2(h-1)\frac{[2\kappa_1\kappa_3 + (3h-5)\kappa_2^2]}{2\kappa_1^4} \\
&\quad + h^2(h-1)\frac{[(7h-11)\kappa_1^2\kappa_4 + 4(h-2)(7h-12)\kappa_1\kappa_2\kappa_3 + 2(h-2)(7h^2-30h+32)\kappa_2^3]}{12\kappa_1^6} \\
&\quad + O(n^{-4}), \\
\mu_3' = \kappa_3' &= \frac{h^3[\kappa_1\kappa_3 + 3(h-1)\kappa_2^2]}{\kappa_1^4} \\
&\quad + h^3(h-1)\frac{[3\kappa_1^2\kappa_4 + 3(7h-10)\kappa_1\kappa_2\kappa_3 + (17h^2-55h+44)\kappa_2^3]}{2\kappa_1^6} + O(n^{-4}), \\
\mu_4' &= \frac{h^4(h-1)}{4(\kappa_1^8)}\begin{bmatrix} 8\kappa_1^3\kappa_5 + 12(7h-9)\kappa_1^2\kappa_3^2 + 2(59h-79)\kappa_1^2\kappa_2\kappa_4 \\ + 4(149h^2-442h+327)\kappa_1\kappa_2\kappa_3 \\ + (339h^3-1649h^2+2673h-1447)\kappa_2^4 \end{bmatrix} + O(n^{-5}).
\end{aligned}\right\} \quad \ldots\ldots(5)$$

If we put $h=1-\frac{\kappa_1\kappa_3}{3\kappa_2^2}$ the leading term of μ_3' vanishes. We now find for the first moments and cumulants of y

$$\left.\begin{aligned}
\text{Mean}=\kappa_1'&=1-h(1-h)\frac{\kappa_2}{2\kappa_1^2}+h(1-h)(2-h)(1-3h)\frac{\kappa_2^2}{8\kappa_1^4}+O(n^{-3}),\\
\mu_2'=\kappa_2'&=h^2\frac{\kappa_2}{\kappa_1^2}-h^2(1-h)(1-3h)\frac{\kappa_2^2}{2\kappa_1^4}+O(n^{-3}),\\
\mu_3'=\kappa_3'&=h^3(1-h)\frac{[2(23-49h+23h^2)\kappa_2^3-3\kappa_1^2\kappa_4]}{2\kappa_1^6}+O(n^{-4}),\\
\mu_4'&=3h^4\frac{\kappa_2^2}{\kappa_1^4}+h^4\frac{[3\kappa_1\kappa_4-(19-29h)\kappa_2\kappa_3]}{3\kappa_1^5}+O(n^{-4}),\\
\kappa_4'&=\frac{h^4}{3\kappa_1^5}[3\kappa_1\kappa_4-4(4-5h)\kappa_2\kappa_3]+O(n^{-4}).
\end{aligned}\right\}\quad\ldots\ldots(6)$$

Since h, κ_1, κ_2, κ_3 are not independent, these expressions can clearly be written in many ways. However the above are probably the simplest. Since the standard deviation of y is

$$\frac{h\kappa_2^{\frac{1}{2}}}{\kappa_1}\left[1-\frac{(1-h)(1-3h)\kappa_2}{4\kappa_1^2}\right]+O(n^{-\frac{5}{2}})$$

we may conveniently say that

$$\xi=\left[\left(\frac{x}{\kappa_1}\right)^h+\frac{h(1-h)\kappa_2}{2\kappa_1^2}\left\{1-\frac{(2-h)(1-3h)\kappa_2}{4\kappa_1^2}\right\}-1\right]\frac{\kappa_1}{h\kappa_2^{\frac{1}{2}}}\left[1+\frac{(1-h)(1-3h)\kappa_2}{4\kappa_1^2}\right],\quad\ldots\ldots(7)$$

is almost normally distributed with mean zero and standard deviation unity. On substituting a given value of x in this equation we can evaluate ξ and thus obtain the probability of as large or a larger deviation of x from κ_1 by means of Sheppard's table. This transformation will be referred to as transformation *A*. For large values of n, and particularly when h does not differ greatly from $\frac{1}{3}$, it will be sufficient to put

$$\xi=\left[\left(\frac{x}{\kappa_1}\right)^h+\frac{h(1-h)\kappa_2}{2\kappa_1^2}-1\right]\frac{\kappa_1}{h\kappa_2^{\frac{1}{2}}}.$$

It follows that the first two measures of deviation of ξ or y from normality are

$$\begin{aligned}
\gamma_1'&=\frac{(1-h)[2(23-49h+23h^2)\kappa_2^3-3\kappa_1^2\kappa_4]}{2\kappa_1^3\kappa_2^{\frac{3}{2}}}+O(n^{-\frac{3}{2}})\\
&=\frac{\kappa_3}{54\kappa_1^2\kappa_2^{\frac{3}{2}}}(9\kappa_1\kappa_2^2\kappa_3+46\kappa_1^2\kappa_3^2-27\kappa_1^2\kappa_2\kappa_4-54\kappa_2^3)+O(n^{-\frac{3}{2}}),\\
\gamma_2'&=\frac{3\kappa_1\kappa_4-4(4-5h)\kappa_2\kappa_3}{3\kappa_1\kappa_2^2}+O(n^{-2})\\
&=\frac{9\kappa_1\kappa_2\kappa_4-20\kappa_1\kappa_3^2+12\kappa_2^2\kappa_3}{9\kappa_1\kappa_2^3}+O(n^{-2}).
\end{aligned}$$

Thus γ_1' is of order $n^{-\frac{3}{2}}$, as compared with $\gamma_1 = \frac{\kappa_3}{\kappa_2^{\frac{3}{2}}}$ of order $n^{-\frac{1}{2}}$, and

$$\frac{\gamma_2'}{\gamma_2} = 1 - \frac{4(4-5h)\kappa_2\kappa_3}{3\kappa_1\kappa_4} + O(n^{-1})$$

$$= 1 - \frac{4\kappa_3(5\kappa_1\kappa_3 - 3\kappa_2^2)}{9\kappa_1\kappa_2\kappa_4} + O(n^{-1}).$$

This may be small numerically Thus for the χ^2 distribution $\gamma_2'/\gamma_2 = -\frac{1}{27}$. When it is not numerically small, a further transformation will reduce it to the order of n^{-1}.

ADJUSTMENT OF THE FOURTH MOMENT

In order that γ_2' should tend to zero with n^{-2}, κ_4' must tend to zero with n^{-4}. This can be secured by adding a suitable constant to x, so that κ_1 assumes any desired value g, whilst the other cumulants are unchanged. In order that the leading term of κ_4' should vanish, we must make $3g\kappa_4 - 4(4-5h)\kappa_2\kappa_3 = 0$. Hence

$$g = \frac{12\kappa_2^2\kappa_3}{20\kappa_3^2 - 9\kappa_2\kappa_4}, \quad h = \frac{16\kappa_3^2 - 9\kappa_2\kappa_4}{20\kappa_3^2 - 9\kappa_2\kappa_4},$$

whilst for x we must substitute $x + g - \kappa_1$. It is clear that we can now remove our restriction on the order of κ_1 when n is large. On the other hand if κ_2 and κ_3 are of order n, κ_4 must be of the same order. We can now state the following theorem:

"If the variable x be so distributed that its cumulants κ_2, κ_3, κ_4 tend to infinity with r, and no later cumulant tends to infinity more rapidly than n^{r-4}, then if

$$g = \frac{12\kappa_2^2\kappa_3}{20\kappa_3^2 - 9\kappa_2\kappa_4}, \qquad h = \frac{16\kappa_3^2 - 9\kappa_2\kappa_4}{20\kappa_3^2 - 9\kappa_2\kappa_4},$$

and

$$z = \left(1 + \frac{x - \kappa_1}{g}\right)^h,$$

z is so distributed that its third and fourth cumulants tend to zero with n^{-3} and n^{-4} respectively."

As above, the theorem remains true if x is multiplied by any constant; but it breaks down if $20\kappa_3^2 = 9\kappa_2\kappa_4$ or $16\kappa_3^2 = 9\kappa_2\kappa_4$. It follows that for the distribution of z, $\gamma_1'' = O(n^{-\frac{3}{2}})$ and $\gamma_2'' = O(n^{-2})$.

To prove the theorem we substitute g for κ_1 in equations (5), and give h its appropriate value. This is equivalent to putting

$$g\kappa_3 = 3(1-h)\kappa_2^2,$$

$$g^2\kappa_4 = 4(1-h)(5-4h)\kappa_2^3.$$

The cumulants of z are most simply expressed in terms of

$$b = \frac{h}{g} = \frac{16\kappa_3^2 - 9\kappa_2\kappa_4}{12\kappa_2^2\kappa_3}, \qquad c = \frac{1-h}{g} = \frac{\kappa_3}{3\kappa_2^2}.$$

They are

$$\left.\begin{aligned} \text{Mean} = \kappa_1'' &= 1 - \tfrac{1}{2}bc\kappa_2\left[1 + \tfrac{1}{4}(b+2c)(2b-c)\kappa_2\right] + O(n^{-3}), \\ \mu_2'' = \kappa_2'' &= b^2\kappa_2\left[1 + \tfrac{1}{2}c(2b-c)\kappa_2\right] + O(n^{-3}), \\ \mu_3'' = \mu_3'' &= -b^3c(3b^2 - 3bc + c^2)\kappa_2^3 + O(n^{-4}), \\ \kappa_4'' &= b^4c(12b^3 + 2b^2c - 138bc^2 + 247c^3)\kappa_2^4 - 2b^4c + O(n^{-5}). \end{aligned}\right\} \quad \ldots(8)$$

For purposes of calculation it is convenient to write $h = bg$, so that

$$\xi = \left[\left(1 + \frac{x-\kappa_1}{g}\right)^{bg} + \tfrac{1}{2}bc\kappa_2\left\{1 + \tfrac{1}{4}(b+2c)(2b-c)\right\} - 1\right]\left[\frac{1 - \frac{1}{4}c(2b-c)\kappa_2}{b\kappa_2^{\frac{1}{2}}}\right], \quad \ldots(9)$$

or less accurately

$$\xi = \frac{\left(1 + \dfrac{x-\kappa_1}{g}\right)^{bg} + \tfrac{1}{2}bc\kappa_2 - 1}{b\kappa_2^{\frac{1}{2}}},$$

is almost normally distributed about zero with unit standard deviation and measures of deviation from normality

$$\gamma_1'' = -c(3b^2 - 3bc + c^2)\kappa_2^{\frac{3}{2}} + O(n^{-\frac{3}{2}})$$

$$= \frac{-592\kappa_3^4 + 756\kappa_2\kappa_3^2\kappa_4 - 243\kappa_2^2\kappa_4^2}{432\kappa_2^{\frac{3}{2}}\kappa_3} + O(n^{-\frac{3}{2}}),$$

$$\gamma_2'' = c(12b^3 + 2b^2c - 138bc^2 + 247c^3)\kappa_2^3 - \frac{2c\kappa_5}{\kappa_2^2} + O(n^{-3})$$

$$= \frac{5908\kappa_3^6 - 16344\kappa_2\kappa_3^4\kappa_4 + 11826\kappa_2^2\kappa_3^2\kappa_4^2 + 2187\kappa_2^3\kappa_4^3 - 864\kappa_2^2\kappa_3^3\kappa_5}{1296\kappa_2^6\kappa_3^2} + O(n^{-3}).$$

This transformation will be referred to as transformation B. It may give γ_1 larger or smaller than transformation A. It will certainly give γ_2 smaller for large values of n. However we shall see that in some cases, even when n is as large as 100, transformation A may give a better approximation. h is negative if

$$20\kappa_3^2 > 9\kappa_2\kappa_4 > 16\kappa_3^2.$$

In this case approximation (9) is clearly invalid for negative or very small positive values of $1 + \dfrac{x-\kappa_1}{g}$. It may, however, be found that for sufficiently large values of n such values are very improbable. As will be seen later, in the case of χ^2 this expression cannot become negative, and (in the theory of large samples) cannot vanish.

It is clear that other transformations are possible. If we desired to obtain a highly symmetrical distribution without concerning ourselves with its flatness, we could either choose g so that the term of κ_5' vanished, or we could reduce γ_1' to the order $n^{-\frac{3}{2}}$ by making the leading term in the expression (6) for μ_3' vanish, that is to say by choosing g so that

$$g^2(27\kappa_2\kappa_4 - 46\kappa_3^2) - 9g\kappa_2^2\kappa_3 + 54\kappa_2^3\kappa_4 = 0.$$

Or for a given value of n we could obtain κ_3' and κ_4' in equations (5) as functions of g and h, and choose g and h so that both vanish. For small values of n this would involve the calculation of many terms in the expansions of κ_3' and κ_4' and would rarely be worth doing. It will be seen that in some cases transformation A is fully sufficient for practical purposes.

APPLICATION TO THE χ^2 DISTRIBUTION

For the χ^2 distribution $\kappa_1 = n$, $\kappa_2 = 2n$, $\kappa_3 = 8n$, $\kappa_4 = 48n$, $\kappa_5 = 384n$. Transformation A is Wilson and Hilferty's transformation, whose very satisfactory fit is shown in their paper. In transformation B,

$$g = \frac{12n}{13}, \quad b = \frac{5}{12n}, \quad c = \frac{2}{3n}, \quad h = \frac{5}{13}.$$

So

$$\xi'' = \left[\left(\frac{13x^2 - n}{12n}\right)^{\frac{5}{13}} + \frac{5}{18n}\left(1 + \frac{7}{48n}\right) - 1\right]\frac{6\sqrt{2n}}{5}\left(1 - \frac{1}{18n}\right) \quad \ldots\ldots(10)$$

is almost normally distributed about zero with unit standard deviation, and

$$\gamma_1'' = \frac{19}{27(2n)^{\frac{3}{2}}} + O(n^{-\frac{5}{2}}),$$

$$\gamma_2'' = \frac{35}{18n^2} + O(n^{-3}).$$

Comparing these with the values found above for transformation A we find

$$\frac{\gamma_1''}{\gamma_1'} = \frac{19}{32} + O(n^{-1}),$$

$$\frac{\gamma_2''}{\gamma_2'} = \frac{35}{4n} + O(n^{-2}).$$

Thus the symmetry is always improved, but for $n < 9$, γ_2 is increased. However, for large values of n the fit is considerably improved. Thus for $n = 10$ we find, for different values of P, the values of χ^2 given in Table I.

TABLE I

Approximations to the χ^2 distribution

P	0·8	0·5	0·2	0·05	0·01
True value of χ^2	6·179	9·342	13·442	18·307	23·209
Approximation A	6·191	9·349	13·419	18·298	23·246
„ B	6·182	9·339	13·430	18·288	23·194

or less accurately,

$$\xi' = \left[\frac{\left(\dfrac{x}{np}\right)^{\frac{p+2q}{3q}} - 1}{p+2q} + \frac{p-q}{18npq}\right] 3\sqrt{(npq)},$$

is almost normally distributed about zero with unit standard deviation, and with

$$\gamma_1' = \frac{-(p-q)^2(17+2p)}{54(npq)^{\frac{3}{2}}} + O(n^{-\frac{3}{2}}),$$

$$\gamma_2' = \frac{-11p^2 - 8pq + q^2}{9npq} + O(n^{-2}).$$

$|\gamma_1'|$ is always less than $|\gamma_1|$ γ_2' vanishes when $p = \frac{1}{2}(3\sqrt{(3)} - 5) = \cdot 0981$ and does not exceed $\frac{1}{9}\gamma_2'$ for lower values of p. Also $|\gamma_2'| < |\gamma_2|$ if $p < \frac{1}{26}(16 - \sqrt{(126)})$, or $\cdot 1875$. So for small values of p the transformation is likely to be valuable, and even in the neighbourhood of $\frac{1}{2}$, γ_2' has the moderate value of $2n^{-1}$. If $p < \frac{1}{2}$ we should perform transformation A on $n - x$. The transformation has a very slight effect in the neighbourhood of $p = \frac{1}{2}$, but here the normal approximation is already very close.

Transformation B does not seem to be well suited to the binomial distribution, since both p^2q^2 and $(p-q)^2$ occur in the denominator of γ_2'', and numerical tests show that it is not so efficient.

A few numerical examples, given in Table II, show the degree of accuracy

TABLE II

Approximations to P in the binomial distribution

	True value	Normal	"A"	Gram-Charlier
$p=0\cdot25$, $n=10$, $x=5\cdot5$	0·9803	0·9851	0·9798	0·9731
$p=0\cdot10$, $n=100$, $x=3\cdot5$	0·0078	0·0157	0·0085	0·0094
$p=0\cdot10$, $n=100$, $x=17\cdot5$	0·9900	0·9938	0·9900	0·9877

reached by transformation A. In each case we take a value of x intermediate between two integers. The second column gives the probability that x should exceed this value. The third column gives the probability that a normal deviate should exceed $\dfrac{x-np}{\sqrt{(npq)}}$. The third column gives the probability that a normal deviate should exceed ξ' calculated from the shorter form of equation (11). The last column gives the same calculation derived from the first two terms of the Gram-Charlier series for the probability when the binomial deviate is expanded

in a series of Hermitian polynomials. The correction to the probability of column 2 is $+\frac{p-q}{6\sqrt{(npq)}}e^{-\frac{1}{2}\xi^2}$, where $\xi=\frac{x-np}{\sqrt{(npq)}}$. It will be seen that in each case transformation A gives the best fit, though further terms in the Gram-Charlier series would doubtless give a still better fit.

Discussion

The approximations here given must obviously be used with caution. Like the normal approximation to the binomial, they can give absurd results. For example finite probabilities are found for negative values of χ^2, though even for $n=5$ the probability of a negative value is less than 10^{-5} according to equation (3), whilst equation (10) holds down to a probability of 2×10^{-4}. When the transformations are applied to any particular distribution it is always possible that series (4) may converge very slowly even for fairly large values of n. This is so for example, in the case of the estimate of β_2. The values of this estimate in the samples of n from a normal population are distributed round 3 with cumulants of order n. But their numerical coefficients increase so rapidly that, on applying transformation B,

$$h=\frac{13}{5},\quad g=\frac{4n}{5},$$

and γ_1'' is greater than γ_1 until n exceeds 110.

The main object of this investigation has been to make it possible to deal with the χ^2 distribution and that of Fisher's (1935) u statistics when expectations are small. This will be done in separate papers.

REFERENCES

FISHER, R. A. (1928). "Moments and product-moments of sampling distributions." *Proc. London Math. Soc.* XXX, 199–238.

—— (1935). "The detection of linkage with recessive abnormalities." *Ann. Eugen., London,* VI, 339–51.

HALDANE, J. B. S. (1937). "The exact value of the moments of the distribution of χ^2, used as a test of goodness of fit, when expectations are small." *Biometrika*, XXIX, 133–43.

"STUDENT" (1908). "The probable error of a mean." *Biometrika*, VI, 1–25.

WILSON, E. B. and HILFERTY, M. M. (1931). "The distribution of Chi-square." *Proc. nat. Acad. Sci. Washington*, XVII, 684.

DE WINTON, D. and HALDANE, J. B. S. (1935). "The genetics of *Primula sinensis*. Linkage in the diploid." *J. Genet.* XXXI, 67–100.

MISCELLANEA

A Note on a Form of Tchebycheff's Theorem for Two Variables.

By P. O. BERGE

University of Stockholm

LET $f(x, y)$ be a frequency function with the means

$$\bar{x}=0, \quad \bar{y}=0,$$

and the moments of the second order σ_1^2, σ_2^2, μ_{11}.

We shall try to find a limiting value for the probability, that $|x| < k\sigma_1$, and $|y| < k\sigma_2$, i.e. that the point (x, y) lies within the rectangle R, limited by the straight lines

$$x = \pm k\sigma_1, \quad y = \pm k\sigma_2.$$

We shall use a procedure similar to that used by Prof. Karl Pearson in his paper "On Generalized Tchebycheff Theorems in the mathematical theory of statistics" (*Biometrika*, XII, 284) but confining our attention to the case where only the moments of the first and second order are given.

The function
$$\frac{1}{k^2(1-t^2)}\left[\frac{x^2}{\sigma_1^2} - 2t\frac{xy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2}\right] \qquad \text{......(1)}$$

is, if $t^2 \leqslant 1$, not negative and outside the rectangle R, greater than 1.

We get

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{k^2(1-t^2)}\left[\frac{x^2}{\sigma_1^2} - 2t\frac{xy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2}\right]f(x,y)\,dx\,dy = \frac{2(1-tr)}{k^2(1-t^2)} \geqslant 1 - \iint_{(R)} f(x,y)\,dx\,dy,$$

where the last integral is to be taken over the region inside of the rectangle R, and r is the correlation-coefficient.

Thus
$$\iint_{(R)} f(x,y)\,dx\,dy \geqslant 1 - \frac{2(1-tr)}{k^2(1-t^2)}. \qquad \text{......(2)}$$

The expression on the left is the probability that

$$|x| < k\sigma_1 \quad \text{and} \quad |y| < k\sigma_2,$$

and our problem will now be to find a minimum value for

$$\frac{2(1-tr)}{k^2(1-t^2)}.$$

Differentiating this expression with regard to t and equating to zero, we get

$$-r(1-t^2) + 2t(1-tr) = 0$$

or

$$t = \frac{1}{r}(1 \pm \sqrt{1-r^2}).$$

But t must satisfy the condition $t^2 \leqslant 1$. We find that this condition is fulfilled if we choose

$$t = \frac{1}{r}(1 - \sqrt{1-r^2}). \qquad \text{......(3)}$$

For this value of t we get

$$\frac{2(1-tr)}{k^2(1-t^2)} = \frac{1+\sqrt{1-r^2}}{k^2}.$$

Hence we see that

the probability that $|x| < k\sigma_1$ and $|y| < k\sigma_2$ is greater than or equal to

$$1 - \frac{1+\sqrt{1-r^2}}{k^2}. \qquad \text{......(4)}$$

It is possible to show that the value (4) is the most accurate we can reach, when only the moments of the first and second order are known. Consider a discontinuous frequency function with, (i) the probability

$$\frac{1+\sqrt{1-r^2}}{4k^2}$$

attached to each of the four points

$$\left(k\sigma_1,\quad \frac{k\sigma_2(1-\sqrt{1-r^2})}{r}\right),\quad \left(\frac{k\sigma_1(1-\sqrt{1-r^2})}{r},\quad k\sigma_2\right),$$

$$\left(-\frac{k\sigma_1(1-\sqrt{1-r^2})}{r},\quad -k\sigma_2\right),\quad \left(-k\sigma_1,\quad -\frac{k\sigma_2(1-\sqrt{1-r^2})}{r}\right),$$

—these four points lying on the rectangle R—and (ii) the probability

$$1-\frac{1+\sqrt{1-r^2}}{k^2}$$

attached to origin.

If $\bar{x}'$ and $\bar{y}'$ are the means and $\sigma_1'^2$, $\sigma_2'^2$, μ_{11}' are the moments of the second order of the above-mentioned discontinuous frequency function, we get

$$\bar{x}'=\bar{y}'=0,$$

$$\sigma_1'^2=k^2\sigma_1^2\frac{1+\sqrt{1-r^2}}{2k^2}+k^2\sigma_1^2\frac{(1-\sqrt{1-r^2})^2(1+\sqrt{1-r^2})}{2k^2r^2}=\sigma_1^2,$$

$$\sigma_2'^2=k^2\sigma_2^2\frac{(1-\sqrt{1-r^2})^2}{r^2}\frac{(1+\sqrt{1-r^2})}{2k^2}+k^2\sigma_2^2\frac{1+\sqrt{1-r^2}}{2k^2}=\sigma_2^2,$$

$$\mu_{11}'=4k\sigma_1\frac{k\sigma_2(1-\sqrt{1-r^2})}{r}\;\frac{1+\sqrt{1-r^2}}{4k^2}=r\sigma_1\sigma_2.$$

This frequency function consequently has its means at the origin; its two standard deviations are equal to σ_1 and σ_2 and the correlation-coefficient is equal to r.

This example shows that the value (4) is the best one in the sense, that it is impossible to say that the probability for

$$|x|<k\sigma_1 \quad\text{and}\quad |y|<k\sigma_2$$

is *greater* than

$$1-\frac{1+\sqrt{1-r^2}}{k^2}$$

unless we know higher moments.

Finally, it is of interest to compare certain numerical values of the limiting expression (4) with known values of the integral within the rectangle, R, in the case where the frequency function $f(xy)$ is the Normal Bivariate Surface

$$\phi(x,y)=\frac{1}{2\pi\sqrt{1-r^2}}e^{-\frac{1}{2(1-r^2)}(x^2+y^2-2rxy)}$$

k	$\int_{-k}^{k}\int_{-k}^{k}\phi(x,y)\,dxdy$			$1-\frac{1+\sqrt{1-r^2}}{k^2}$		
	$r=0$	$r=0{\cdot}5$	$r=1{\cdot}0$	$r=0$	$r=0{\cdot}5$	$r=1{\cdot}0$
2·0	0·9111	0·9171	0·9545	0·5000	0·5335	0·7500
2·6	0·9814	0·9823	0·9907	0·7041	0·7240	0·8521

$r=1{\cdot}0$ corresponds to the case of *one* variable.

While the approximation is more accurate for high values of r, it will be seen that for this particular frequency surface the limits given by my theorem are drawn very conservatively.

Note on J. B. S. Haldane's Paper: "The Exact Value of the Moments of the Distribution of χ^2." (Biometrika, XXIX, 133–143.)

By W. G. COCHRAN

In this paper Haldane points out (p. 142) a difference between his results for the mean and variance of χ^2 in a $2 \times n$-fold contingency table when the expectation p is fixed and the results obtained by me in my paper (*Annals of Eugenics*, Vol. VII, part III, p. 211). The difference is that I have $(n-1)$ throughout where Haldane has n. Haldane writes, "my own results would appear to be slightly more accurate than Cochran's", which might, I think, give the impression that both Haldane's results and mine are only approximations. In fact, both results are mathematically exact, the difference between them being one of definition of χ^2. My paper is almost entirely concerned with the distribution of χ^2 when the expectation p is not known. In the results which I gave for the distribution of χ^2 when p is known, I retained the term $S(x-\bar{x})^2$ in the numerator of χ^2 instead of $S(x-np)^2$, to facilitate comparison between this and my other results. Thus my χ^2 has $(n-1)$ degrees of freedom, whereas Haldane's χ^2 has n degrees of freedom.

Unfortunately I did not emphasize this point in the passage concerned, and as it may have appeared misleading to others besides Haldane, I welcome this opportunity of drawing attention to it. Haldane's χ^2 is, of course, the one which is normally appropriate in testing the departure from independence when the expectation is known.

The reader may perhaps also wonder why in § 3 I have replaced $\frac{\bar{x}}{n}(n-\bar{x})$ by κ_2 in the denominator of χ^2. This was done because I am discussing the distribution of χ^2 in arrays in which the mean of the sample is equal to the mean of the population, so that in this paragraph the two expressions can be regarded as equivalent.